

Language universals and linguistic complexity

Three case studies in core argument marking

Kaius Sinnemäki

*Academic dissertation to be publicly discussed, by due permission of the Faculty of Arts
at the University of Helsinki in Arppeanum (auditorium, Snellmaninkatu 3), on the 19th
of October, 2011 at 12 o'clock.*

General Linguistics
Department of Modern Languages
University of Helsinki
2011

Copyright © 2011 Kaius Sinnemäki

The articles in Part II have been included in the paperback version with permission from their respective publishers.

ISBN 978-952-10-7259-8 (PDF) – <http://ethesis.helsinki.fi>

Unigrafia Oy

Helsinki 2011

Abstract

In this dissertation I study language complexity from a typological perspective. Since the structuralist era, it has been assumed that local complexity differences in languages are balanced out in cross-linguistic comparisons and that complexity is not affected by the geopolitical or sociocultural aspects of the speech community. However, these assumptions have seldom been studied systematically from a typological point of view.

My objective is to define complexity so that it is possible to compare it across languages and to approach its variation with the methods of quantitative typology. My main empirical research questions are: i) does language complexity vary in any systematic way in local domains, and ii) can language complexity be affected by the geographical or social environment? These questions are studied in three articles, whose findings are summarized in the introduction to the dissertation.

In order to enable cross-language comparison, I measure complexity as the description length of the regularities in an entity; I separate it from difficulty, focus on local instead of global complexity, and break it up into different types. This approach helps avoid the problems that plagued earlier metrics of language complexity.

My approach to grammar is functional-typological in nature, and the theoretical framework is *basic linguistic theory*. I delimit the empirical research functionally to the marking of core arguments (the basic participants in the sentence). I assess the distributions of complexity in this domain with multifactorial statistical methods and use different sampling strategies, implementing, for instance, the Greenbergian view of universals as diachronic laws of type preference. My data come from large and balanced samples (up to approximately 850 languages), drawn mainly from reference grammars.

The results suggest that various significant trends occur in the marking of core arguments in regard to complexity and that complexity in this domain correlates with population size. These results provide evidence that linguistic patterns interact among themselves in terms of complexity, that language structure adapts to the social environment, and that there may be cognitive mechanisms that limit complexity locally. My approach to complexity and language universals can therefore be successfully applied to empirical data and may serve as a model for further research in these areas.

Acknowledgments

My initial connection to linguistics took place in college when my teacher of Finnish language and literature taught us the terms morphology, syntax, semantics, and so on. On the examination, I could not provide the right definitions for some of those terms. Evidently, I was not too interested in the subject. That state of affairs continued until I made the right connections. In that process, Chris Pekka Wilde and Richard Brewis played key roles; I gratefully acknowledge their input in making me enthusiastic about linguistics and showing its meaningful applications.

I have had the privilege of having two outstanding supervisors for my research, Professor Fred Karlsson and Professor Matti Miestamo. I am deeply indebted to each for their continual interest, encouragement, and many helpful comments on my work. Through them, I have learned the meticulous work required for scientific writing and have been empowered to work my way through a topic that at first seemed unattainable for a Ph.D. student.

I am very grateful to my preliminary examiner and opponent, Dr. Michael Cysouw, for truly encouraging feedback and constructive criticism. I also express my deep gratitude to Dr. Guy Deutscher for a thorough preliminary review that helped me situate my work in a wider scholarly context.

Over the years, I have greatly benefited from discussions with colleagues in Finland and abroad, in matters of academic advice, friendship, or collegial fellowship. Although all cannot be named here, I must gratefully mention Eija Aho, Anu Airola, Antti Arppe, Harald Baayen, Dik Bakker, Balthasar Bickel, Östen Dahl, Casper de Groot, Anders Enkvist, August Fenk, Gertraud Fenk-Oczlon, David Gil, Jeff Good, Ekaterina Gruzdeva, Riho Grünthal, Tom Güldemann, Harald Hammarström, Martin Haspelmath, Kari Hiltula, Tommi Jantunen, Patrick Juola, Kirsti Kamppuri, Mari-Sisko Khadgi, Don Killian, Seppo Kittilä, Kimmo Koskeniemi, Ritva Laury, Miroslav Lehecka, Jaakko Leino, Frank Lichtenberg, Ivan Lowe, Annu Marttila, John McWhorter, Urho Määttä, Jussi Niemi, Jyrki Niemi, Esko Niiranen, Urpo Niskanen, Alexandre Nikolaev, Santeri Palviainen, Irina Piippo, Kari Pitkänen, Pekka Posio, Johanna Ratia, Jouni Rostila, Jack Rueter, Geoffrey Sampson, Niels Smit, Andrew

D.M. Smith, Mickael Suominen, Lauri Tarkkonen, Johanna Vaattovaara, Martti Vainio, Ulla Vanhatalo, Liisa Vilkki, Max Wahlström, Stefan Werner, Hanna Westerlund, Juha Yiniemi, Anssi Yli-Jyrä, Jussi Ylikoski, and Jan-Ola Östman.

I would also like to thank the Linguistics section of the University of Helsinki for an outstanding working environment and my colleagues in Langnet—the Finnish Graduate School for Language Studies—, 2006-2009, as well as my colleagues in the Helsinki Circle for Typology and Field Linguistics, and those on the board of the Linguistic Association of Finland, 2004-2008, for many memorable moments. Surprisingly, some of the most joyful academic times I have spent have been as an administrator, serving with my colleagues of the Steering Committee and Executive Board of Langnet, 2007-2008; their cheerful fellowship is a source of great pleasure. A typologist could not do much without data. I would like to thank Janet Bateman, Mark Donohue, Aone van Engelenhoven, Eva Lindström, Laura McPherson, and Lourens de Vries for providing important data on their language of expertise.

My research has been made possible economically by several financiers: Langnet, the Academy of Finland, the University of Helsinki, the Finnish Cultural Foundation, and the Emil Aaltonen Foundation. Their financial support is gratefully acknowledged. I am also grateful to Bernard Comrie and Martin Haspelmath for arranging a stay at the Max Planck Institute for Evolutionary Anthropology in Leipzig, as well as to the Free Evangelical Church of Vihti and to Manor Hotel Ruurikkala (Youth with a Mission Ruurikkala) for arranging facilities for my remote work.

I would also like to thank my parents, Pekka and Soile Sinnemäki, my sister, Annukka Sinnemäki, and her family, as well as my relatives, in-laws, and friends, for their support during the long years that went into the preparation of this dissertation.

Finally, the journey to complete this work would not have been possible without the constant support of my Heavenly Father. I agree with Solomon: “It is the Lord who gives wisdom; from him come knowledge and understanding” (Proverbs 2:6).

Above all, I would like to thank my beautiful, beloved wife Laura and our lively, treasured son Samuel for patience and awesome support, and all the life, joy, love, and beauty they bring to my life. I wholeheartedly dedicate this work to them; I am forever thankful to God for you.

Contents

Abstract.....	iii
Acknowledgments	iv
Contents	vi
Abbreviations.....	viii

PART I Introduction

Chapter 1 Overview	1
1.1. Objectives and research questions.....	2
1.2. Theoretical approach, methods, and data	3
1.3. List of articles included	4
1.4. Main results	6
1.5. Structure of the dissertation.....	6
Chapter 2 Theoretical and methodological issues.....	8
2.1. The current work in its typological context	8
2.2. Background to typological research on language complexity	10
2.3. What is complexity?.....	13
2.3.1. Complexity vs. difficulty	13
2.3.2. Local vs. global complexity	16
2.3.3. Measuring complexity	20
2.3.4. Types of complexity	22
2.3.5. Complexity and rarity	28
2.3.6. Relationship between complexity and some linguistic notions	29
2.4. Defining the domain of inquiry	32
2.5. Complexity in the case studies	36
2.5.1. Article 1: Complexity trade-offs in core argument marking	36
2.5.2. Article 2: Complexity in core argument marking and population size	39
2.5.3. Article 3: Word order in zero-marking languages	40
2.5.4. Evaluation	42

Chapter 3 Sampling, statistical methods, and data.....	43
3.1. Evaluation of language universals.....	43
3.1.1. On sampling.....	43
3.1.2. Statistical methods.....	48
3.2. Data.....	49
Chapter 4 Results and discussion.....	52
4.1. Complexity in typology.....	52
4.2. Evaluation of language universals.....	54
4.3. Empirical results.....	55
4.3.1. Question 1: Systematic complexity variation.....	56
4.3.2. Question 2: Complexity in its geographical and sociocultural context	60
4.3.3. Relevance of different complexity types to language universals.....	62
4.4. On explanations.....	63
Chapter 5 Conclusion.....	66
5.1. Main scholarly contribution.....	66
5.2. Issues for further research.....	68
References.....	72

PART II Articles

List of Articles.....	93
ARTICLE 1: Complexity trade-offs in core argument marking.....	94
Commentary on Article 1.....	117
ARTICLE 2: Complexity in core argument marking and population size.....	123
Commentary on Article 2.....	143
ARTICLE 3: Word order in zero-marking languages.....	148
Commentary on Article 3.....	193

Abbreviations

1	first person
3	third person
ABS	absolutive
AOR	aorist
CAUS	causative
DAT	dative
DEF	definite
DIR	directional
DIST	distal demonstrative
DU	dual
ERG	ergative
FOC/TNS	particle of focus/tense
HAVE	verb forming prefix meaning ‘have N’
IMPOSS	impossibility
IND	indicative
INDF	indefinite
M	masculine
MUT	mutation
NEG	negative
OBJ	object
PFV	perfective
PL	plural
PRET	preterite
PST	past
RED	reduplicated
RL	realis
SBJ	subject
SG	singular
yi	unknown meaning (Guirardello 1999: 62-70)

PART I Introduction

Chapter 1 Overview

This dissertation is about language complexity. The approach taken here is typological, meaning that the aim is to develop methods for a cross-linguistic study of complexity in order to determine how complexity varies across languages and provide explanations for that variation. The main title, **Language universals and linguistic complexity**, reflects this overall theme, while the subtitle, **Three case studies in core argument marking**, specifies the grammatical locus and the practical manner of realizing the theme.

Until recently, complexity has not been widely researched in linguistics, despite its growing importance in other disciplines. Whatever discussion there has been has mostly centered on the repetition of two fundamental assumptions. On the one hand, it has been assumed that while different languages may vary as to the locus of complexity, for instance, one having complex morphology and another having many word order rules, there is a balancing out (or trade-off/compensation) of these differences in typological comparison (e.g., Hockett 1958: 180-181; Bickerton 1995: 35, 76; Crystal 1997: 6; Aboh and Smith 2009: 4). On the other hand, it has been assumed that language complexity has nothing to do with its geographical or social environment (e.g., Sapir 1921: 219; Kaye 1989: 48). The usual implication, or companion, of these claims is that all languages are at an approximately equal level of complexity overall (an assertion henceforth referred to as **the equi-complexity hypothesis**).

Despite the centrality of these claims to many branches of linguistics, especially to structuralism, generativism, and creolistics, their validity has rarely been subjected to systematic cross-linguistic investigation. The outcome has been their dogmatization and the ensuing lack of empirical and theoretical research on language complexity. In addition, the little cross-linguistic research that has been done has mostly focused on complexity in phonology or morphology, leaving other grammatical variables largely untouched (e.g., Kusters 2003; Shosted 2006). As a result, we have little idea of the general scope of the alleged balancing effects or the role of the geographical and/or sociocultural environment in complexity variation. This is a rather sorry state for assumptions that have been central to theoretical linguistics (see Sampson 2009). These

obvious gaps in the research on language complexity are directly related to the aims and research questions in this dissertation.

1.1. Objectives and research questions

My objective in this dissertation is two-fold. First, I intend to define complexity in a way that is amenable to cross-linguistic comparison and to measurement in different grammatical domains. Second, my aim is to gain deeper insights into the cross-linguistic variation of grammatical complexity by approaching it from the viewpoint of quantitative typology. This means using large and well-balanced samples, controlling confounding factors, and using statistical methods for testing hypotheses. The main empirical research questions are as follows:

- Is there any systematic cross-linguistic variation in the grammatical complexity of languages in a particular domain?
- Can grammatical complexity be affected by the social environment of a speech community, for instance, by population size?

Instead of searching for complexity trade-offs, I am assuming that any correlations are equally interesting from a typological perspective. I assume that there are no reasons why complexity could not be one parameter along which typological variables could vary systematically among themselves or vis-à-vis other anthropological variables, because this is characteristic of linguistic structures in general (Bickel 2007: 240). Given these research questions, my aim is not to determine whether all languages are equally complex; rather, I argue that, methodologically, this question may be completely unattainable.

1.2. Theoretical approach, methods, and data

The approach taken here could be described as functional-typological. I acknowledge that grammatical structure may be affected, although not necessarily dictated, by its function or its frequency of use, for example (see Givón 1979, 2001; Haspelmath 2008). Typological distribution of complexity may thus be fruitfully explained in terms of functional motivations, most notably in terms of the general principles of economy and distinctiveness. Moreover, these principles may be built into the complexity metric itself, as has been done here (see Article 2). Yet, because numerous factors can affect complexity distributions, I argue that multiple causation is needed to explain them.

The theoretical framework adopted here is known as *basic linguistic theory* (usually abbreviated as BLT; see Dixon 1997: 128-138, 2009, 2010; Dryer 2006). BLT could be characterized as a cumulative, informal, and framework-neutral approach to describe and analyze grammatical phenomena. It draws mostly on traditional grammar, structuralism, and early generative work and, informed by analyses and comparisons of different languages over many years, consists of concepts that have been of lasting value, while being open to the incorporation of new ideas. BLT suits the purpose of quantitative typology very well, as it provides easy coding of typological variables.

From the outset I assume that typological distributions are best characterized as the probabilistic outcome of the different forces that shape language structure, not as absolute, hard-wired constraints (see Dryer 1998; Maslova 2000; Bickel 2010). From this viewpoint, typology is not merely the flipside of the Chomskyan Universal Grammar, attempting to discover absolute constraints on possible human language, but a sub-discipline of its own, with its own research agenda, questions, and methods that focus on cross-linguistic diversity (Bickel 2007; Nichols 2007). My interest is thus not in the limits of grammatical complexity as much as in the probabilistic distribution of complexity and the possible correlations that involve complexity.

I further endorse a multi-methodological approach to the study of language universals. Since there is no consensus in the field as to how such things as the effect of areas should be controlled for, I model areas in multiple ways. While in most of my case studies universals are approached from a synchronic point of view, I also adopt the

Greenbergian view of language universals as diachronic laws of type preference (e.g., Greenberg 1978) as implemented in the Family Bias Theory of Bickel (2008b, 2011). To study the research questions, I use large and well-balanced samples (varying from 50 to approximately 850 languages), draw data (mostly) from reference grammars, and use statistical methods, such as multiple logistic regression, to assess whether the typological distributions of complexity are statistically meaningful and independent of confounding factors.

I limit the study of the research questions to one particular functional domain, namely that of core argument marking. A functional domain in Givón's (1981) sense is a group of closely related functions encoded at least by some languages (e.g., the passive, aspect) (Miestamo 2007: 293). In core argument marking, three types of morphosyntactic marking—head marking (or agreement), dependent marking (or case marking), and rigid word order—interact in distinguishing “who is doing what to whom.” Focusing on coding strategies in the same functional domain enables the study of cross-linguistically comparable variables whose relationship is also theoretically well-motivated.

As for the notion of complexity, I tie it to a more general framework of complexity (Rescher 1998) and keep it separate from difficulty of language use (e.g., Dahl 2004; see Chapter 2). I further argue that language complexity can be fruitfully measured when the focus is on particular types of complexity in their local contexts.

1.3. List of articles included

My dissertation consists of this introduction (Part I) and the three articles listed below (Part II):

Article 1: Kaius Sinnemäki 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*. Studies in Language Companion Series 94, 67-88. Amsterdam/Philadelphia: John Benjamins.

Article 2: Kaius Sinnemäki 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*. Studies in the Evolution of Language 13, 126-140. Oxford: Oxford University Press.

Article 3: Kaius Sinnemäki 2010. Word order in zero-marking languages. *Studies in Language* 34(4): 869-912.

In these case studies, or articles, the two research questions introduced above are broken down into three, more particular questions.

- 1) Is there a systematic cross-linguistic balancing of complexity between head marking, dependent marking, and rigid word order? (Article 1)
- 2) Is complexity of core argument marking related to sociocultural properties of a speech community, namely, population size? (Article 2)
- 3) Does morphological simplicity (the absence of morphological strategies) correlate with Subject-Verb-Object (SVO) word order, which has been claimed to be the most economical linear order available for argument discrimination? (Article 3)

Article 1 is the thematic and chronological starting point for the work and focuses on the interaction among the strategies. The two other articles take up more specific themes, focusing on particular aspects of the complexity of core argument marking. The weight is on complexity variation within core argument marking (Articles 1 and 3), while the interaction of complexity with the geographical and sociocultural environment receives slightly less attention (Article 2, and to some extent, Article 3).

1.4. Main results

The main results of this dissertation are of three kinds. First, I argue for the relevance and usefulness of the notion of complexity to functional-typological research. I do so a) by providing a detailed discussion of the notion of complexity and how it could be applied to language (focusing on a particular type of complexity in its local context), b) by showing how the complexity of a morphosyntactic domain can be approached typologically, and c) by arguing how some general principles, such as economy and distinctiveness, can be used to measure complexity. This approach clarifies the notion of language complexity and has proved to be useful when applied to empirical data; hopefully, it may foster further typological research on language complexity.

Second, the empirical results suggest that various kinds of significant trends occur in core argument marking in terms of complexity. Some trade-offs exist (Article 1), but positive correlations (Article 3, Commentary on Article 1) and correlations between complexity and population size (Article 2) are possible as well. Differences also exist as to the type of complexity that is meaningful to the correlations. These results provide evidence for cognitive mechanisms that may limit particular types of complexity (Miestamo 2008: 31-32), but they also suggest that language adapts to its context and is not reducible to biology (e.g., Givón 2009).

Third, the current work is one of the first attempts at implementing multiple logistic regression and, more generally, a multi-methodological approach to modeling language universals. Although the applicability of logistic regression is limited in typology, this work illustrates its suitability for modeling typological distributions and, together with Bickel (2008b) and Cysouw (2010), may serve as a model for future research.

1.5. Structure of the dissertation

This dissertation consists of two parts. Part I introduces the research topic, summarizes the main findings, and ties the case studies to current typological discussion. This part consists of five chapters. The current chapter outlines the research questions, the main

results, and the structure of the work. Chapter 2 presents the theoretical background of the typological approach adopted here as well as the notion of language complexity and defines the domain of inquiry. Chapter 3 discusses sampling, the statistical evaluation of language universals, and the data. Chapter 4 presents and discusses the main results and evaluates their scholarly contribution. Chapter 5 summarizes the main idea of this study and discusses possibilities for further research. Part II consists of the articles, that is, the case studies, in the order given in Section 1.3. Each article is followed by a short commentary and, where necessary, errata.

Chapter 2 Theoretical and methodological issues

In this chapter I discuss theoretical and methodological issues related to the typological study of complexity. In Section 2.1, I describe the typological approach used, and in Section 2.2, the background to complexity in typology. Section 2.3 discusses language complexity, Section 2.4 the domain of inquiry, namely, core argument marking, and Section 2.5 the ways in which complexity was measured in the case studies.

2.1. The current work in its typological context

In this study, I approach language complexity from a typological perspective. Typology is here understood as a (sub)discipline of its own within linguistics, not as an alternative methodology to the generativist goal of determining what is a possible human language (Bickel 2007: 239-240, 248). Typology is concerned with uncovering cross-linguistic diversity, making generalizations based on data from a wide range of languages, and studying interrelationships among linguistic patterns and interactions with other anthropological patterns, such as social, cognitive, and genetic patterns. The major contribution of typology, this study included, is within the broad range of other anthropological undertakings, not limited to Cognitive Sciences (narrowly defined), as is Chomskyan generative grammar (cf. Chomsky 1995; Ritter 2005).

As a discipline, typology may be informally divided into three main streams: qualitative, quantitative, and theoretical typology (Bickel 2007). I use this informal grouping here to help locate my dissertation in its immediate typological context.

Most work that comes under the rubric of typology is qualitative in nature. The purpose, then, is to develop variables for capturing similarities and differences in and across languages, and applying and refining them with new data. Insightful cross-linguistic work often begins by comparing two very different types of languages; only later is the work applied to a larger sample (Nichols 2007: 233-234). Even work based on one language can be described as (qualitative) typological in nature, as long as the research questions approach language from a general typological perspective.

Quantitative typology is about researching statistical trends in the distribution of variables established by qualitative typology (that is, studies of language universals), and about developing and using statistical methods and modeling techniques suitable for such data (that is, methodological studies). Examples of research that is mostly of the former kind include Dryer's (1992) research on word order correlations and Nichols' (1992) study of morphological patterns, while studies that are largely of the latter type include Maslova's (2000) mathematical approach to universals and Janssen, Bickel, and Zúñiga's (2006) discussion of statistical methods.

Lastly, theoretical typology is a matter of elucidating the internal structure of particular grammatical domains as well as of explaining the typological distribution of grammatical variables. Explanations involve economy and iconicity (Haiman 1983), processing preferences (Hawkins 2004), frequency (Haspelmath 2008), principles of language change (Croft 2000), and population history (Nichols 1992).

Most typological work cross-cuts two or more of these three main streams, and my study is no exception. My research is qualitative in nature in its attempt to define language complexity as well as the coding strategies interacting in core argument marking in a way that is cross-linguistically comparable (Chapter 2). It is quantitative in nature in that it uses large and stratified samples, studies typological distributions with statistical methods, and conducts reliability tests for the results, in the spirit of Janssen, Bickel, and Zúñiga (2006) (Chapter 3). My work is also related to theoretical typology in its attempt to provide explanations for the attested universals (Section 4.4). Since sampling techniques, large samples (up to approximately 850 languages) and statistical tests play an important role in my dissertation (especially in Article 3), this work is very much quantitative in nature and thus belongs to a small minority within typology, which, however, is probably its best-known segment outside the field (cf. Nichols 2007: 232-235).

As for the earlier typological work in Finland, it has mostly been qualitative and/or theoretical in nature. This applies largely to the typological works by Esa Itkonen, Seppo Kittilä, and Matti Miestamo. Quantitative methods have rarely been used and when they have, their use has been mostly limited to issues of sampling (e.g., Miestamo 2005), seldom involving statistical tests, such as correlation tests (e.g.,

Miestamo 2009), or multidimensional scaling (e.g., Vilkki, forthcoming). The current study therefore is the first attempt in Finland to pursue quantitative methods in typology in a serious way.

2.2. Background to typological research on language complexity

Complexity is and has been a controversial concept in linguistics, being simultaneously friend and foe. On the one hand, complexity is a central notion in linguistic theorizing and description. Leafing through an introductory textbook or “essential readings” in almost any subfield, one can hardly avoid encountering the notion, and not only in an informal way, but also as being more or less central to the subfield or theory at large. For example, notions such as markedness in naturalness theory (e.g., Dressler et al. 1987), recursion in generative grammar (e.g., Hauser, Chomsky, and Fitch 2002), and economy in functional linguistics (e.g., Haiman 1983) are all related to complexity (see Section 2.3.6). Many grammatical terms are also based on complexity; for instance, the grouping of verbs into intransitive, transitive, and ditransitive is based on the number of arguments that the verbs take, reflecting a difference in the complexity of argument structure.

Moreover, little has been done on language complexity in general or on its typological variation in particular. Even though complexity is often used in the literature as an important theoretical notion, its definition is, unfortunately, often vague. Comrie (1992), for one, discusses the development of complexity in languages, but purposefully leaves the definition of complexity open. In addition, despite decades of work since Greenberg’s (1966) seminal work, typological research on these issues began only in the last ten years, including McWhorter (2001) and its commentary in *Linguistic Typology* 5(2/3), Kusters (2003), Dahl (2004), Hawkins (2004), Trudgill (2004) and its commentary in *Linguistic Typology* 8(3), Maddieson (2006), Miestamo (2006), McWhorter (2007), Bane (2008), Miestamo, Sinnemäki, and Karlsson (2008), Pellegrino et al. (eds.) (2009), Sampson, Gil, and Trudgill (2009), Mackenzie (2009), Lupyán and Dale (2010), Good (2010), and Bakker et al. (2011). To my knowledge, only five antecedents occurred in the 1990s, namely, Comrie (1992), Nichols (1992),

Perkins (1992), Juola (1998), and Fenk-Oczlon and Fenk (1999). However, once the empirical cross-linguistic research on language complexity really began, it has at times generated heated opposition (e.g., DeGraff 2001).

This controversy may come as a surprise to an outsider to linguistics, but there are historical and methodological reasons behind it. I will not expound on the historical sides of the issue, as they have been thoroughly treated elsewhere (e.g., Kusters 2003: 1-5; Newmeyer 1986: 39ff; Sampson 2009). Here I will provide only brief comments.

In sum, the historical reasons for the lack of comparative research into complexity derive from the structuralist withdrawal from the earlier racist equation of language complexity with the degree of development of a certain people or nation. Although this withdrawal was and is definitely justified, it left linguistics with two tenets that, practically speaking, caused the pendulum to swing to the opposite direction and stopped comparative research on language complexity for several decades. One tenet is that language structure has nothing to do with its geographical or sociocultural setting; the other is that, although complexity may vary across languages in certain domains, the differences are balanced out when compared cross-linguistically, meaning that all languages are approximately equally complex overall (cf. Chapter 1). The recent research on language complexity acknowledges that the structuralist withdrawal from the earlier racist ideas was fully justified, but maintains that typological research into complexity is warranted and may increase our understanding of language and its relationship to other anthropological variables (e.g., McWhorter 2001; Kusters 2003; Shosted 2006; Lupyan and Dale 2010). My dissertation continues this line of thinking. I withdraw from any value judgments concerning language complexity, but maintain that it is possible to conduct scholarly work on language complexity in an ethically sound way.

As for methodology, structuralists largely rejected the comparison of language complexity as unfruitful.¹ Later critics have judged the approximation of overall language complexity as problematic, criticized the attempted metrics as superficial, and

1 The reasons for this rejection are unclear, although the structuralist ideal of describing each language in its own terms may have seemed to be an insurmountable obstacle for comparative work. This ideal, however, is not a real problem for cross-linguistic comparison (see Haspelmath 2010).

lamented that the metrics lack psychological plausibility (e.g., DeGraff 2001). The criticism is partly justified, especially as it concerns the estimation of the overall complexity of languages (cf. Miestamo 2008; Deutscher 2009; also Section 2.3.2).

The accusation of superficiality, however, reflects merely a difference in theoretical approach.² Most work on language complexity has been done from a non-generative approach, which generally takes language use and linguistic diversity seriously and avoids concepts far removed from surface patterns. Yet most of the criticism has come from the generative approach, whose proponents are not necessarily very interested in linguistic diversity, but rather in abstract general principles that are thought to lay hidden beneath the surface patterns (e.g., Boeckx 2009). This criticism thus reflects the deeply ingrained differences between the generative and the non-generative approaches in many of their basic assumptions about language (see Evans and Levinson 2009) and is reflected in such things as the lack of cross-referencing between these two approaches. Compare, for instance, the references in Frank (2004), a paper on grammatical complexity from a generative approach, to the references in my dissertation.

Meanwhile, the accusation that complexity metrics lack psychological plausibility, on the other hand, is in danger of repeating earlier problems evident in Chomsky's evaluation measure (see Section 2.3.1). One way to overcome this is precisely by keeping complexity and difficulty apart and study the processing responses of different complexity metrics independently (e.g., Dahl 2004; Miestamo 2008; Hawkins 2009).

Despite these controversies (or maybe even because of them), the typological research into complexity has increased markedly in the last decade, so my dissertation is surely not the first to try and answer the questions laid out in Chapter 1. However, the research of the last decade focused on such things as the interaction between phonological and morphological variables, between the number of utterance elements (phonemes, syllables; Fenk-Oczlon and Fenk 1999, 2008), or between syllable inventory size and the degree of inflectional synthesis (Shosted 2006). My study is one

2 This is obvious in Aboh and Smith's (2009: 8) critique of my work on complexity (Article 1).

of the first attempts to bring in at least one syntactic variable (see also Parkvall 2008; Miestamo 2009).

The correlation between complexity and sociocultural environment has drawn increased attention as well. Perkins (1992) argued for an inverse relationship between the complexity of language's deictic system and cultural complexity. McWhorter (2001, 2007) has argued that language contact has simplified creoles as well as languages of wider communication (e.g., English). Kusters (2003) argued that language contact simplifies verbal inflection, while Hay and Bauer (2007) argued for a correlation between phoneme inventory size and population size. In addition, Lupyan and Dale (2010), using data from *The World Atlas of Language Structures* (henceforth the *WALS*; Haspelmath et al. 2005), argued that morphosyntactic complexity correlated inversely with population size. My work continues this line of research and focuses on complexity in one functional domain and correlates it with population size.

2.3. What is complexity?

In this section, I describe my general approach to language complexity. I discuss the separation of complexity from difficulty (Section 2.3.1), the focus on local instead of global complexity (Section 2.3.2), the measuring of complexity (Section 2.3.3), the types of complexity (Section 2.3.4), and the relationship of complexity to frequency (Section 2.3.5), along with a few other important notions in linguistics (Section 2.3.6).

2.3.1. Complexity vs. difficulty

I begin the discussion by considering the relation of complexity to difficulty. According to the *Oxford Advanced Learner's Dictionary* (Hornby and Wehmeier 2007), the adjective *complex* has two senses: i) consisting of many interrelated parts and ii) being difficult to understand. These two senses emerge in the theoretical discussion of complexity as well. On the one hand, it has been argued, both in the natural sciences and in the recent debate on language complexity, that complexity always depends on our model of reality, on the theoretical framework, and ultimately on the observer (e.g.,

Popper 1959; Gell-Mann 1995; Simon 1996; Kusters 2003, 2008; Bower 2009). Edmonds (1999: 50) goes so far as to claim that complexity is primarily a matter of our model of reality and is only projected into reality via the model. Thus, complexity has no ontological status of its own. In this view, complexity is subjective and could be broadly described as emphasizing the difficulty of understanding.

This view is not unanimously shared, however. Rescher (1998: 16-21), for one, argues that complexity is a real and general property of real world elements, whose complexities exist regardless of whether anyone observes them. At the same time, our best **practical index** of complexity is the difficulty of coming to cognitive terms with it, that is, the amount of effort spent on its description. Therefore, it is our **description** of complexity, and not the ontological properties of real world elements, that depend on models or theories. In this view, there is no reason to assume that the difficulty of describing a phenomenon would create complexity or project it to reality, but that this difficulty reflects true complexity, to the extent, of course, that the description is a good model of reality. As a result, the general notion of complexity is not purely a matter of real world elements or the limits of our cognitive capacity, but involves both aspects, since only through our limited cognitive capacities can we gain access to reality.

In my dissertation, I follow Rescher's (1998) view, because it provides the most general approach to complexity. While it recognizes the subjective nature of complexity metrics, this view shows the relationship of the epistemic side of complexity to its ontological side. In addition, this view sheds light on why different scholars may arrive at different results when studying the complexity of an entity: this state of affairs may simply reflect the fact that different models (to the extent they are good models of reality) describe different aspects of reality, and thus, they capture different aspects of the complexity as well. A linguistic example related to this issue is treated in Section 2.5.1, where I discuss the varying and practically opposing opinions regarding the complexity of word order. In effect, no single model or approach can capture the full complexity of a real world entity, because the range of facts about an entity is inexhaustible (Rescher 1998) and because we may need multiple lenses with which to view the notion of complexity itself (Page 2011).

The difficulty of describing a phenomenon is also very different from the difficulty of its use. Miestamo (2008) introduces two terms for describing these, namely, absolute and relative complexity. **Absolute complexity** is a matter of the number of parts and interrelations in a system, whereas **relative complexity** is a matter of the cost or difficulty of using or processing a certain grammatical construction, for instance. While Kusters (2008) treats both types of phenomena as examples of relative complexity, the former relative to a theory and the latter to a user, there are at least four reasons why it is better to keep these two strictly separate (see also Dahl 2004).

First, description and operation are two separate tasks, which can be performed independently of one another. Native speakers talk fluently without thinking about language description, while, to some degree, description is possible without fluency in the target language (e.g., via bilingual informants).

Second, if a general approach to language complexity is based on the difficulty of use, then there is the problem of finding a user-type neutral definition for complexity (Miestamo 2008: 24-29). The point is that the relative difficulty of different grammatical phenomena varies among different user-types, namely, among speakers, hearers, first-language acquirers, and second-language learners (Kusters 2003). One phenomenon is easy for speakers and first-language acquirers, but difficult for hearers and second-language learners, while another phenomenon may be easy to all user-types except second-language learners (e.g., Kusters 2003: 45-62). To avoid this problem, a general approach to complexity is best done from a more objective (or theory-based) perspective.

Third, keeping complexity separate from difficulty helps avoid the problems that plagued the evaluation measure of early generative grammar (Chomsky 1965; Chomsky and Halle 1968).³ The evaluation measure was used for choosing among competing

3 In Naturalness Theory, a fundamental assumption is that naturalness judgments are grounded in extralinguistic reality, that is, in the cognitive and anatomical bases of language as well as in the ease vs. the difficulty of language production and comprehension (Dressler et al. 1987: 11-12; Mayerthaler 1987: 26-27; Dressler 2003). Therefore, naturalness is explicitly a theory about the difficulty of use, and it faces similar problems as those encountered by descriptive length in early generative grammar.

theories the one that most closely resembled the way children acquire language – a vital step in advancing the framework from descriptive to explanatory adequacy. It was assumed that the framework that provided the shortest description of the system would also provide the closest link with the ease/difficulty of language acquisition.

However, this assumption encountered many problems, including the lack of a non-arbitrary basis for the selection of alternative theoretical accounts (Prideaux 1970) and the remark that the shortest description was not necessarily the most plausible one psychologically (Kiparsky 1968). Calls for the psychological plausibility of complexity metrics are still heard (e.g., DeGraff 2001), but I maintain here that the best way to avoid repeating earlier errors in this domain is to keep complexity separate from cost or difficulty.⁴

This leads directly to the fourth reason: when the two concepts of complexity and difficulty are treated separately, it is possible to determine independently the processing responses of different types of complexity (see Hawkins 2004, 2009).⁵ Such comparison may show that some types of complexity have stronger processing responses than others, but this is only to be expected and should in fact caution us to avoid strong *a priori* evaluation of the plausibility of different metrics.

Having separated complexity from difficulty and having emphasized the need to approach complexity from an absolute/objective/theory-oriented view, I continue by separating local complexity from global complexity.

2.3.2. Local vs. global complexity

As has often been observed, the concept of complexity is difficult to define. **In general, complexity may be characterized as the number and variety of elements and the elaborateness of their interrelational structure** (Rescher 1998: 1; Simon 1996: 183-

4 The evaluation measure was used to compare different theoretical accounts of one and the same phenomenon, while description length in the current complexity debate is about describing a structure within a particular theoretical framework, not across frameworks (Miestamo 2008: 28).

5 The need to separate complexity (structural or syntactic) from difficulty (cognitive or processing complexity) is also evident in Croft and Cruse (2004: 175) as well as in Givón (2009: 11-14).

184; Hübler 2007: 10). In linguistics, a general intuition is that “more structural units/rules/representations mean more complexity” (Hawkins 2009: 252). But when it comes to operationalizing complexity for actual measurement, it soon becomes clear that no unified definition exists. Scholars have proposed numerous ways to measure complexity: Edmonds (1999: 136-163) identifies forty-eight different formulations, used mostly in natural and social sciences (e.g., algorithmic information complexity, entropy, and minimum size), while Lloyd (2001) lists around forty formulations in his inventory. Complexity as a general, overall notion thus seems to escape unified and precise verbal formulae.

This leads directly to an important terminological distinction, which is crucial in discussing complexity, namely, the distinction between local and global complexity (Edmonds 1999; Miestamo 2006, 2008). **Local complexity** is about the complexity of some part of an entity, while **global complexity** is about the overall complexity of that entity. As I have already intimated, there are problems in measuring the global complexity of language and thus also in evaluating the equi-complexity hypothesis.

There are at least four issues in connection with these problems (see Miestamo 2006, 2008, and Deutscher 2009 for fuller accounts). First, no typological complexity metric can take into account all relevant aspects of a language’s grammar, because it is simply beyond the capacities of a single linguist or even the community of linguists to produce a comprehensive description of the grammar of any language. Miestamo (2006, 2008) calls this **the problem of representativity**. The crux of the problem is not merely a practical one, the limitation of the labor force, but also the limitations of human knowledge (Rescher (1998: 25ff). The number of descriptive facts about real world elements is unlimited, and, therefore, our knowledge of the world will always remain incomplete.⁶ The only instance where the attainable level of representativity might suffice is when the complexity differences are very clear, as seems to be the case in McWhorter’s (2001) and Parkvall’s (2008) comparison of creoles with non-creoles.⁷

6 See also Moscoso del Prado Martín (2010). Based on analyses of text corpora, he claims that the effective complexity (see Section 2.3.3 for the definition) of languages is practically unlimited.

7 Note that even if one opposed Parkvall’s (2008) measure of global complexity, creoles seem to form a distinct typological class in light of cross-linguistic data (see Bakker et al. 2011).

Second, there is no principled way of comparing various aspects of complexity to one another or evaluating their contribution to global complexity. For example, how should morphological and syntactic complexity be weighed, and how much do they contribute to the global complexity of a language? Miestamo (2006, 2008) calls this **the problem of comparability**. Again, when the differences are clear and all or most of the criteria point in the same direction, it might be possible to compare global complexity, for instance, by comparing two closely related languages (Dahl 2009).

The third point is a result of the two previous points. Although it appears to be possible to compare the global complexity of languages when the differences are clear, it is not possible to make these comparisons when differences are more subtle or when different criteria contradict each other. This leads to the following conclusion: it is possible to evaluate the equi-complexity hypothesis only as an exceptionless, absolute universal, and the hypothesis seems to have been refuted by the demonstration that some languages, such as creoles or closely related languages, differ from other languages in terms of (approximate) global complexity (see McWhorter 2001, 2007; Parkvall 2008; Dahl 2009; Bakker et al. 2011).

However, the attempt to test whether there is a general statistical tendency to limit the global complexity of languages encounters insurmountable methodological problems, owing to the issues discussed above. Comprehensive complexity metrics, such as those proposed by Nichols (2009), provide interesting estimates, but since these metrics assume equal weights for complexities in different domains, it is unclear how accurately they approximate the global complexity of languages. What this means is that, even if some languages were shown to differ in terms of approximate global complexity, it appears to be impossible to determine whether such tests have any bearing on the equi-complexity hypothesis as a possible statistical tendency. One possible way to overcome the problem of complexity weighing is to scrutinize grammatical structures in untaged texts in mathematical ways (e.g., Juola 1998, 2008; Moscoso del Prado Martín 2011).

Yet while these methods reveal complexity trade-offs, they are unable to capture the global grammatical complexity of languages. One reason is that they cannot capture all information concerning word order phenomena, because in untaged texts, word

order regularities can be detected only by noting multiple instances of lexical collocations of the same lexemes in similar or different orders, and this is insufficient for noting all word order regularities (Miestamo 2008: 28). In addition, texts are merely the output of the grammatical system, and, as such, they can provide only an indirect view of the complexity of that system.

Fourth, the general picture that emerges from the supporters of the equi-complexity hypothesis is that equal complexity of languages requires complexity trade-offs to be a general principle in language (e.g., Hockett 1958; Bickerton 1995). If this were true, then one could at least disprove the hypothesis as a statistical universal by examining the presence or absence of possible trade-offs, or negative correlations, in a handful of feature pairs (cf. Shosted 2006; Maddieson 2006). However, this assumption seems premature, since positive correlations are not in conflict with general balancing effects (Fenk-Oczlon and Fenk 2008). Preliminary computer simulations further suggest that it is possible that only a fraction of negative complexity correlations are significant, even when global complexity is held constant. This result indicates that trade-offs are only indirectly related to the equi-complexity hypothesis (Sinnemäki, in preparation). Correlations among a limited set of features may thus tell very little about the global complexity of languages, suggesting that the relationship between complexity trade-offs and the equi-complexity hypothesis is indirect at best and unfalsifiable at worst.

Based on these issues, I find it methodologically impossible to answer reliably whether the equi-complexity hypothesis is a statistical universal or not. I further concur with McWhorter (2001: 134) in that, even though it would be possible to rank languages along some complexity scale, it is unclear what the intellectual benefit of such an endeavor would be. Much more promising is the study of the local complexity of languages. This has been advocated by several linguists (LaPolla 2005; Miestamo 2006, 2008; Deutscher 2009; Good 2010; among others) and seems to be acceptable even to the critics of language complexity research (e.g., Ansaldi and Matthews 2007: 6).

2.3.3. Measuring complexity

As argued in Section 2.3.1, it is more feasible to approach complexity from an objective or theory-oriented viewpoint than from a subjective or user-related viewpoint. It is sometimes claimed that there is no consensus among linguists as to how to define objective complexity (e.g., Ansaldo and Nordhoff 2009). In this section, I describe how I define complexity and how it can be related to a more general framework of complexity, drawing especially from Rescher (1998), Dahl (2004), and Miestamo (2008: 24-29). I further argue that behind the terminological differences, a marked consensus exists among many linguists as to the criteria for complexity.⁸

In Section 2.3.2, I defined the general notion of complexity as the number and variety of elements and the elaborateness of their interrelational structure. This general notion can be made more widely applicable by drawing from the principles of information theory (beginning with Shannon 1948). In algorithmic information theory, a well-known measure of complexity is **Kolmogorov complexity**, which measures the description length needed to specify an object (e.g., Li and Vitányi 2008; also Chaitin 1987). It has been argued by many linguists that description length can fruitfully be applied to measuring language complexity as well: the longer the description of a linguistic structure, the more complex it is (e.g., Dahl 2004; Bane 2008; Juola 2008; Miestamo 2008; Moscoso del Prado Martín 2010). For instance, it requires a shorter description to account for the morphological structure of the verb in Maybrat, where the only inflection on the verb is the person prefix (1), than in Turkish, where several morphemes can occur on the verb at the same time, including a person affix (2).

8 An underlying consensus seems to exist even more generally in the sciences of complexity, owing to interrelations between the many definitions for complexity. Lloyd (2001), for one, proposes a simple three-way typology for complexity: 1. difficulty of description, 2. difficulty of creation, 3. degree of organization. These types can be further situated in Rescher's (1998) more general approach to complexity under descriptive, generative, and structural modes, respectively (see Section 2.3.4).

Maybrat (North-Central Bird's Head; Dol 1999: 69)

- (1) *Fane y-tien.*
 pig 3M-sleep
 'The boar sleeps.'

Turkish (Turkic; Wurzel 2001: 380)

- (2) *dol-dur-ma-yabil-ir-di-m*
 fill(itself)-CAUS-NEG-IMPOSS-AOR-PRET-1SG
 'I could have refrained from filling (it/something) in'

In applying Kolmogorov complexity to particular problems, a given piece of text or description is often compressed by using a computer algorithm, such as a zip-program. The idea is that the shorter the output of the algorithm, the less complex is the object. Although compression algorithms have been used in earlier studies of language complexity (e.g., Juola 1998, 2008; Bane 2008), I follow instead Miestamo (2008: 24-25) and adopt the idea of description length on a much more general level, where it is more useful to apply descriptive tools developed by linguists than those by mathematicians (cf. Vulcanovic 2007). The complexity of structures is also compared at a level on which clear differences can be found, for instance when comparing the complexity of verbal morphology in Maybrat and Turkish.

What I mean by "clear differences" can be simply reduced to the presence vs. absence of overt coding. In other words, a language which has a case marking of the object is more complex than one that has no object case marking (but only with respect to the case marking of the object, not in general). This corresponds with the use of the term markedness as overt coding (see Section 2.3.6), but following Haspelmath (2006), I use the terms overt vs. non-overt marking instead of markedness and connect these terms with the more general notion of complexity. Although not all linguists would agree that overt vs. non-overt coding reflects a difference in complexity (McWhorter 2001: 145), I believe it reflects the most basic kind of complexity difference, namely, zero vs. non-zero complexity (see also Dahl 2004).

Because my focus here is on grammatical complexity (similarly to e.g., McWhorter 2001, Dahl 2004, and Miestamo 2008), the emphasis is on grammatical regulations rather than resources, in the terminology of Dahl (2004, 2008). **Resources** are the possibilities that the system offers to its users, while **regulations** are the constraints and requirements enforced by the system (Dahl 2008: 154). Resources are thus the inventory of such things as morphemes, words, and constructions available to the user, while regulations refer to the requirements imposed on the user when building utterances. In other words, my focus is not on the complexity of the whole object, but on the regularities and patterns in the object. This corresponds to the notion of **effective complexity**, which refers to the description length of the regularities in a system rather than to the description length of the whole system (Gell-Mann 1994, 1995). By focusing on the former I situate complexity between order and disorder, which corresponds to what scholars usually mean by complexity (Huberman and Hogg 1986; Page 2011).⁹ According to Gell-Mann and Lloyd (2004: 388), effective complexity furthermore is “most useful when comparing two entities, at least one of which has a large value of the quantity in question.” In light of this characterization, my choice of measuring complexity (mostly) as overt vs. non-overt coding is fully justified.¹⁰

2.3.4. Types of complexity

How could we make the general notion of complexity more precise? To measure specific kinds of complexities, a suitable conception of different kinds of complexities is needed. For that purpose, I have adopted Rescher’s (1998: 8-16) method of breaking

9 If the focus were on describing the information content in the whole object, then complexity would be equated with the degree of a system’s randomness. However, that is fundamentally counterintuitive, because Shakespeare’s work, for instance, would then be less complex than random gibberish (Gell-Mann 1995: 2).

10 McAllister (2003) criticizes the notion of effective complexity as being non-unique, which means that different researchers can focus on different sets of regularities in the object, and therefore the effective complexity of an object varies among researchers. However, the problem mostly concerns the global complexity of an object or its subpart. Yet if the focus is on a certain type of complexity in a local context, then the problem is less acute.

up the general notion of complexity into different “modes.” The major modes in this taxonomy are the epistemic, ontological, and functional modes of complexity, all of which are broken down into further modes, as described below (see Table 1 for a condensed summary).

Epistemic complexity is concerned with the formulation of complexity. Its most important aspect for my purpose is description length or **descriptive complexity**. **Functional complexity**, on the other hand, is divided into two modes, the operational and nomic modes of complexity. **Operational complexity** is a matter of a “variety of

Epistemic modes	
Formulaic complexity	
1.	Descriptive complexity: length of the account that must be given to provide an adequate description of a given system.
2.	Generative complexity: length of the set of instructions that must be given to provide a recipe for producing a given system.
3.	Computational complexity: amount of time and effort involved in resolving a problem.
Ontological modes	
Compositional complexity	
1.	Constitutional complexity: number of constituent elements (e.g., in terms of the number of phonemes, morphemes, words, or clauses).
2.	Taxonomic complexity (or heterogeneity): variety of constituent elements, that is, the number of different kinds of components (e.g., tense-aspect distinctions, clause types).
Structural complexity	
1.	Organizational complexity: variety of ways of arranging components in different modes of interrelationship (e.g., phonotactic restrictions, variety of distinctive word orders).
2.	Hierarchical complexity: elaborateness of subordination relationships in the modes of inclusion and subsumption (e.g., recursion, intermediate levels in lexical-semantic hierarchies).
Functional complexity	
1.	Operational complexity: variety of modes of operation or types of functioning (e.g., cost-related differences concerning the production and comprehension of utterances).
2.	Nomic complexity: elaborateness and intricacy of the laws governing a phenomenon (e.g., anatomical and neurological constraints on speech production; memory restrictions).

Table 1. Modes of complexity (Rescher 1998: 9; Karlsson, Miestamo, and Sinnemäki 2009: viii-ix).

modes of operation or types of functioning” and is related to such things as the cost-related differences between the production and comprehension of linguistic utterances (in my definition of complexity, this mode is actually better treated under the general notion of cost or difficulty of use, not under the notion of complexity). **Nomic complexity**, on the other hand, is a matter of “elaborateness and intricacy of the laws governing the phenomenon at issue” and is related to the anatomical and neurological constraints on language production and processing (see Rescher 1998: 9-14; Karlsson, Miestamo, and Sinnemäki 2008: viii-ix). **Ontological modes of complexity** are the most important for my purpose, because they characterize the real and objective properties of an entity. Notwithstanding their importance, it is still through the “window” of epistemic complexity that the ontological modes of complexity are measured.

There are also two ontological modes of complexity, namely, compositional and structural complexity. **Compositional complexity** measures the number and variety of constituent elements, that is, constitutional and taxonomic complexity, respectively (Rescher 1998: 9). The more elements there are to a system, the greater is its **constitutional complexity**, while the greater the variety of the system’s elements, the greater its **taxonomic complexity**. In linguistics, these two modes are more commonly known as syntagmatic structure (constitutional complexity) and paradigmatic structure (taxonomic complexity). These modes of complexity can be applied to linguistic form as well as to semantic representation. Constitutional complexity may well be the most widely used aspect of complexity in linguistics. It may be measured as word length in terms of the number of phonemes or syllables (Fenk-Oczlon and Fenk 1999, 2008), as the degree of inflectional synthesis on the verb (Shosted 2006), or as sentence length in terms of the number of clauses (Diessel 2008). Constitutional complexity can also be used to measure the complexity of semantic representation, for instance, the verb’s valency. Taxonomic complexity in linguistics refers to phenomena such as phoneme inventory size (e.g., Shosted 2006; Maddieson 2006), the variety of meanings ascribed to adverbial subordinators (Kortmann 1996), and the number of semantic-pragmatic distinctions that a language makes in a particular domain (e.g., in aspect marking) (McWhorter 2001).

Structural modes of complexity come in two further modes, namely, organizational and hierarchical complexity. **Organizational complexity** is about the number and variety of different modes of interrelationship in which components can be arranged. A linguistic example is the use of different word orders at the phrasal or clausal level, a much researched topic in word-order typology (e.g., Greenberg 1966; Hawkins 1983; Dryer 1992; Cysouw 2010). Another clear example of organizational complexity is the mapping between form and meaning, because that mapping is a matter of the interrelationship between the two (see below).

Hierarchical complexity measures the subordination relationships and their elaborateness in different modes of inclusion and subsumption (Rescher 1998: 9). Linguistic examples are not difficult to find, owing to the centrality of subordination in syntax (e.g., Chomsky 1965; Givón 2009). Recursion, a prime example of hierarchical complexity (cf. Section 2.3.6), has even been claimed to be the most important design feature that separates human language from animal communication (Hauser, Chomsky, and Fitch 2002). Whether that claim is true is a matter of current debate (see Evans and Levinson 2009; van der Hulst 2010), but the centrality of recursion in linguistic theorizing emphasizes how important is the notion of complexity to the field at large.

These modes of complexity are useful in breaking down the general notion of complexity. When they are applied to specific linguistic data, such as rigid order, I assume that a language with rigid order has greater organizational complexity than one without rigid order, but only with respect to this particular feature (rigid order), not in general (see Section 2.5 for a discussion of the complexity of rigid word order). I thus do not hypothesize about how the different types of complexity might contribute to global complexity, but limit my research to particular local contexts. This is the crux of concentrating on specific modes of complexity in their local contexts.

Earlier conceptions of language complexity can be fruitfully related to these modes of complexity (see Table 2). Dahl (2004: 42-46) discusses the notions of system complexity, structural complexity (not to be confused with Rescher's structural modes of complexity), and conceptual complexity. **System complexity** is about "how to express that which can be expressed" (Dahl 2004: 43). In other words, the part "which can be expressed" is about the number of grammaticalized distinctions, and the part

concerned with “how to express” is about the ways in which those distinctions are expressed, that is, how the mapping between meaning and form is done. The former is directly a matter of taxonomic complexity, while the latter is a matter of organizational complexity, since the mapping between meaning and form concerns their interrelationship.

Structural complexity is about the complexity of expressions at some level of description, for instance, the number of morphemes or the amount of hierarchical structure in an expression; it thus crosscuts the compositional and structural modes of complexity (see Table 2). **Conceptual complexity**, on the other hand, is about the length of a concept’s definition, or the number of conceptual conditions in a regulation. For instance, case marking of objects may be limited to animate objects only, so the concept of animacy must be specified in the regulation as a condition for limiting the domain of case marking. Conceptual complexity is a matter of constitutional complexity at the conceptual level, since it specifies the conditions for the occurrence of a specific element.

Dahl’s notion of system complexity is directly related to the principles discussed by Miestamo (2008) for measuring linguistic complexity, namely, **the principle of fewer distinctions** and **the principle of one-meaning–one-form**. The former is about the number of grammaticalized distinctions in a particular functional domain; the latter is about the mapping between meaning and form. These principles can be seen as a way of breaking up Dahl’s notion of system complexity into finer distinctions, that is, into taxonomic and organizational aspects of complexity. Similar principles have also been proposed by Kusters (2003), whose **principle of economy** and **principle of transparency** are roughly identical with Miestamo’s two principles.

Moravcsik and Wirth (1986: 7) employ the terms **syntagmatic complexity** and **paradigmatic complexity**. The former refers to the amount of structure of a linguistic object and is practically identical to Dahl’s notion of structural complexity; the latter refers to the number of subdistinctions within a category, which clearly measures the same thing as Dahl’s notion of system complexity and Miestamo’s principle of fewer distinctions.

The three criteria for complexity used by McWhorter (2007: 21-50) are closely related to the other linguists' conceptions of complexity as well. McWhorter's first criterion, **overspecification**, is concerned with the marking of semantic distinctions, which in most languages is left to context. This criterion is close to Miestamo's principle of fewer distinctions, although it incorporates aspects of typological markedness (cf. Croft 2003: Chapter 4). McWhorter's second criterion, **structural elaboration**, measures the number of rules (e.g., conjugation classes, heterogeneous

Modes of complexity	Different notions of language complexity
Epistemic modes	
• Descriptive complexity	[Absolute complexity (Miestamo 2008); objective complexity (Dahl 2004)].
Ontological modes	
Compositional complexity	
• Constitutional complexity	[Syntagmatic complexity (Moravcsik and Wirth 1986); structural complexity (Dahl 2004)]; conceptual complexity (Dahl 2004).
• Taxonomic complexity	Structural elaboration (McWhorter 2001); [overspecification (McWhorter 2001); economy principle (Kusters 2003); principle of fewer distinctions (Miestamo 2008); paradigmatic complexity (Moravcsik and Wirth (1986); system complexity (Dahl 2004)].
Structural complexity	
• Organizational complexity	[Transparency principle (Kusters 2003); the principle of one-meaning–one-form (Miestamo 2008); system complexity (Dahl 2004); irregularity (McWhorter 2001)]; [structural elaboration (McWhorter 2001); structural complexity (Dahl 2004); syntagmatic complexity (Moravcsik and Wirth 1986)].
• Hierarchical complexity	[Recursion (Hauser, Chomsky, and Fitch 2002); syntactic complexity (Givón 2009)]; [structural complexity (Dahl 2004); syntagmatic complexity (Moravcsik and Wirth 1986)].
Functional modes	
• Operational complexity	[Relative complexity (Kusters 2003); efficiency (Hawkins 2004); naturalness (Dressler et al. 1987; Dressler 2003)].

Table 2. Rescher's (1998) modes of complexity (on the left) contrasted with different notions of language complexity (on the right). Some notions of complexity overlap, because the definitions are limited to the form or function of expressions (e.g., McWhorter's 2001 overspecification and structural elaboration are a matter of different aspects of taxonomic complexity). The brackets around a set of complexity types are for grouping roughly similar definitions of complexity together.

word order) or foundational elements (e.g., phoneme inventory) required to generate surface forms. It thus partly corresponds to Dahl's structural complexity. His third criterion, **irregularity**, is a special type of Dahl's system complexity, because the description of irregularity requires a listing of separate rules for each item.

As is evident from this discussion, there are close interrelations among the different conceptions of linguistic complexity proposed in the literature. In addition, these conceptions may be connected, at least to some extent, to a more general approach to complexity (cf. Rescher 1998). These issues indicate that there is a certain level of consensus among typologists as to the definition of language complexity.

2.3.5. Complexity and rarity

A few words are also in order concerning the possibility of a link between rarity and complexity. It has sometimes been asked whether cross-linguistic rarity is connected to linguistic complexity (e.g., Newmeyer 2007; Harris 2008; Good 2010). Miestamo (2008: 37-38) argues that, on the one hand, no link necessarily exists between rarity and complexity, but on the other hand, a more or less expected correlation exists between rarity and difficulty. This link is related to the observation that structures that are easy and efficient in performance often tend to grammaticalize in languages, and those which are difficult and inefficient in performance tend to grammaticalize more rarely (Hawkins 2004).

However, there seem to be two ways in which rarity and complexity may be related. First, according to Harris (2008), rarities are produced by the coincidence of common historical processes whose coincidence is of low probability. Low probability, in turn, has been associated with complexity (e.g., Edmonds 1999: 148). Second, to the extent that rare patterns require many steps in their evolutionary paths, their evolutionary complexity is greater than that of a pattern which takes fewer historical steps and less time to develop (Dahl 2004: 44, 103-106). Grammatical complexity in general is a product of non-trivial historical processes, which take time to develop or, in Dahl's (2004) terminology, to become mature. Typological rarities may thus be rather

complex grammatically, as is obviously the case in the examples discussed by Harris (2008).

The possible link between rarity and complexity raises the further question of whether languages tend to avoid complexity in general. This is an important question for language complexity research, and I do not intend to provide a definitive answer. A few points, however, should be considered when approaching this question. First, a few scholars have attempted to enumerate the global grammatical complexity of languages (e.g., Parkvall 2008; Nichols 2009). Regardless of the problems related to such measures (see Section 2.3.2), it is worth noting that the global grammatical complexity of languages in these studies formed bell curves. This seems to imply that high degrees of global complexity might be avoided in languages.

Second, my approach, namely, focusing on different types of complexity in their local contexts, suggests that, when answering the above question, it is necessary to examine different types of complexity and different local contexts separately. In this regard consider the following three empirical issues that emerged from my case studies: i) languages rarely use all three coding strategies in core argument marking (Article 1); ii) violation of distinctiveness is rather uncommon in languages (Article 2); and iii) zero marking is typologically rare (Article 3). In the first and second points rarity is related to high complexity, while in the third point it is related to low complexity. This suggests that rarity may be related to complexity in different ways in different situations, and also that one needs to take into account several different examples in order to determine how rarity and complexity are related in general.

2.3.6. Relationship between complexity and some linguistic notions

In Section 2.2, I mentioned that complexity is an important notion in linguistics. In this section, I briefly explicate how complexity is related to many familiar concepts to linguists, such as recursion, markedness, economy, distinctiveness, and transparency.

First, **recursion** in linguistics is the nesting or embedding of a linguistic expression within another expression. For instance, a phrase can be embedded in another phrase of the same kind, as in (3a), where the possessive noun phrase, *our*

neighbor's, has been embedded in a noun phrase *house*. In a similar way, a clause can be embedded in a clause, as in (3b), where the relative clause *who arrived yesterday* has been embedded in the clause *the man is here*. Recursion is thus a clear example of hierarchical relationships, and of hierarchical complexity in language.

- (3) a. *our neighbor's house*
 b. *The man [who arrived yesterday] is here.*

Second, **markedness** refers, technically speaking, to a relationship between two poles of an opposition, one of which is simpler and more general and the other more complex and more constrained (Battistella 1990: 1-2). Linguistically speaking, markedness describes a relationship between entities that have opposing complexity values at a formal or conceptual level – or both. In semantics, markedness refers to the specification of a semantic distinction. For instance, in the opposition in Finnish, *narttu* ‘bitch’ – *koira* ‘dog’, *narttu* is marked, because it refers only to female dogs, while *koira* is unmarked, because it can refer to dogs in general. In terms of complexity, *narttu* has greater conceptual complexity, since its definition requires greater length compared to the definition of *koira* (cf. Dahl 2004: 45-46). In grammar, markedness refers, among other things, to the presence vs. absence of overt coding (e.g., Haspelmath 2006). For instance, in English, the past tense is marked by *-ed*, while the present tense is unmarked: formally, the presence of overt coding is clearly more complex than its absence (e.g., in terms of Dahl’s structural complexity). Based on these characterizations, markedness (as overt coding) is closely related to complexity.

Third, **transparency** refers to a one-to-one mapping between meaning and form (sometimes called the principle of one-meaning–one-form). Departures from that ideal, such as one-to-many mappings or many-to-one mappings, decrease the transparency of mapping between meaning and form. It has frequently been argued that departures from one-to-one mapping between meaning and form increase complexity (see Kusters 2003; Dahl 2004; Miestamo 2008). The reason for this is that when one meaning is expressed by many forms, say, by different allomorphs, the length of describing the mapping of those two allomorphs to their meaning is greater than when the mapping of only one or

two allomorphs has to be described. For instance, compare the allomorphy of the translative case in Finnish to that of the genitive case. The basic meaning of the translative case is a change of state (e.g., *muuttua punaiseksi* ‘turn red’). This case has two allomorphs, namely, *-ksi* and *-kse*. The genitive case, on the other hand, has the allomorph *-n* in the singular and the allomorphs *-e(n)*, *-de(n)*, *-tte(n)*, *-te(n)*, and *-i(n)* in the plural. It is clear that, owing to the greater allomorphy of the genitive case, its description requires greater length than that of the translative case and, in that sense, has greater system complexity in terms of Dahl (2004).

Fourth, according to **distinctiveness**, languages prefer to preserve semantic distinctions through characteristics that are clear at the formal level; in other words, distinctiveness is a preference principle for avoiding many-to-one mappings between meaning and form (e.g., homonymy) and is, therefore, closely related to transparency (see Article 2). Distinctiveness is connected to complexity in at least two ways. First, violation of distinctiveness causes a deviation from one-to-one mapping between meaning and form, which increases the description length of the system at hand (Miestamo 2008). Second, the greater the differences at the formal level in different form-meaning mappings, the greater the complexity required from the system as well. For instance, if a language uses both head and dependent marking in argument discrimination (henceforth called **double marking**), as in West Greenlandic (6), its structural complexity and system complexity are greater than in a language that uses, say, dependent marking only, as in Trumai (4) (see the examples in Section 2.4). Double marking does not violate distinctiveness, but rather enhances it, yet also increases redundancy and thus violates economy (see below). Complexity is therefore one dimension along which it is possible to compare the distinctiveness of different form-meaning pairings.

Finally, **economy** is a preference principle according to which languages prefer economical structures and representations over non-economical ones. It is often related to Zipf’s (1948) principle of least effort, whereby familiar concepts are expressed in shorter forms than non-familiar concepts (see also Haiman 1983: 802; Haspelmath 2008: 5). As discussed in Section 2.3.4, the length of an expression is a direct reflection of its constitutional complexity; it is thus possible to interpret economy as a motivation

for decreasing the constitutional complexity of familiar concepts. Economy is also related to transparency in that a one-to-many mapping between meaning and form violates economy. However, such mappings also increase the system's description length, because how the mapping between meaning and form is done needs to be specified in more detail (Miestamo 2008). These are thus some examples of how economy may be related to complexity.

The purpose of this brief discussion was to illustrate the ways in which some notions familiar to most linguists can be related to complexity. Translating these notions into the same vocabulary may shed more light on them and their interrelationship (cf. Hübler 2007), but more detailed research is a task for the future.

In the following sections, I define the domain of my case studies (Section 2.4) and describe the ways in which complexity was examined (Section 2.5).

2.4. Defining the domain of inquiry

As stated in Chapter 1, my aim is to develop methods for researching the systematic variation of language complexity from a typological perspective. Owing to the immense complexity of language, the study of its complexity variation must be limited. In addition, it is essential to choose the domain of inquiry so that the variables of interest are cross-linguistically comparable. As argued by Miestamo (2006, 2008), a possible *tertium comparationis* for language complexity is provided by functional domains. **Functional domain**, in the spirit of Givón (1981), is a group of closely related semantic and/or pragmatic functions encoded in at least some languages (Miestamo 2007: 293). Examples include the passive, aspect, negation, and possession. Functional domains can be defined at various levels: voice, for instance, can be defined as a functional domain, while the passive can be defined as a subdomain of its own, possibly in a hierarchical relationship to voice. Cross-linguistic comparability can be achieved on other bases as well, as long as the concepts are cross-linguistically comparable (see Haspelmath 2010).

In this study, I focus on the functional domain of core argument marking. **Core argument marking** is about “who is doing what to whom,” that is, discriminating the core arguments of a prototypical two-place transitive verb, one more agent-like (A) and

the other more patient-like (P) (Comrie 2005: 398). Core argument marking is one of the most universal functional domains in languages: almost all languages differentiate between A and P (Palmer 1994).¹¹

This definition limits my study to prototypical two-place transitive verbs, where “prototypical” is meant in language-particular terms. For instance, in English the verb *hit* is a prototypical two-place transitive verb, because it requires two arguments and because these arguments share the same coding with arguments of other prototypical transitive verbs. In Trumai, the verb *fa* ‘hit/kill’ is semantically a two-place verb, but it belongs to a class of idiosyncratic verbs that take a zero-marked subject and a dative object (3a), unlike prototypical transitive verbs in Trumai, which take a zero-marked object and an ergative subject (3b). Guirardello (1999) classifies these idiosyncratic verbs as extended intransitive verbs. Owing to their idiosyncratic behavior, these verbs were excluded from this research, because the focus here is only on the most basic clausal configuration.

Trumai (Isolate; Guirardello 1999: 57, 261)

- (4) a. *Kiki-∅ fa kodechich-ki.*
 man-ABS hit/kill snake-DAT.PL
 ‘The man killed snakes.’
- b. *Karaiw wan-ek pike kapan.*
 non.Indian PL-ERG house make
 ‘The non-Indians made houses.’

In core argument marking, three morphosyntactic coding strategies interact in discriminating the arguments from one another, namely, head marking (or agreement), dependent marking (or case), and word order. The term **coding strategy** is used here following Anderson (2007: 141-148)¹². Other strategies, such as lexical semantics, can also participate in argument discrimination, but they are excluded from this study,

11 Notable exceptions include the Sundic languages Minangkabau and Riau Indonesian which do not seem to have grammaticalized this distinction at all (Gil 2005, 2008).

12 **Formal coding means** is another alternative for this term (see Frajzyngier and Shay 2003).

because my focus is on morphosyntactic strategies.¹³ These three coding strategies operate in other domains as well, for instance, head marking in marking topic continuation, dependent marking in indexing argument properties, and word order in marking information functions. Regardless of their other functions, these coding strategies participate in differentiating the core arguments from one another. Adapting the Systems Interaction principle of Frajzyngier and Shay (2003), I consider these strategies to be independent of one another, but interact in argument discrimination. Independence here means that the coding strategies are “not triggered by any other element occurring in the clause” (Frajzyngier and Shay 2003: 11). In the context of head and dependent marking, this means that these two strategies are not treated as a matter of government, for instance. This choice enables treating the coding strategies separate from one another, which is necessary for measuring their complexity.

I have defined the three coding strategies more precisely in my three articles, so only a brief overview will be given here. Head and dependent marking were defined in the spirit of Nichols (1986, 1992). In **dependent marking**, the arguments occur in distinct forms by virtue of morphological marking, while in **head marking** the form of the verb indicates some properties of the argument(s), such as number. These definitions cover marking by free, bound, and non-linear formatives, that is, by particles, affixes, and tonal or morphophonological alternations (cf. Bickel and Nichols 2007). Dependent marking is done by particles in Tungak (5), by affixes in West Greenlandic (6), and by morphophonological alternations in Nias (7) (/danö/ is the mutated form and /tanö/ the unmutated form). Head marking is done by particles in Tungak (5), by an affix in West Greenlandic (6), and by tones in Korowai (8). A language such as Norwegian (9), without any morphological marking of core arguments, was defined as **zero marking**.

13 Similarly, I have very little to say about the possible complexity differences at the pragmatic level across languages, although that is certainly an interesting topic for future research (e.g., Bisang 2009).

Tungak (Oceanic; Fast 1990: 21)

- (5) *Ri aina ki la ta-taun ani keve pok.*
 PL woman 3PL.SBJ PFV RED-cook OBJ.INDF PL food
 ‘The women cooked the food.’

West Greenlandic (Eskimo-Aleut; Fortescue 1984: 181)

- (6) *Piniartu-p puisi pisar-aa.*
 hunter-ERG seal.ABS catch-3SG.SBJ.3SG.OBJ.IND
 ‘The hunter caught the seal.’

Nias (Sundic; Brown 2001: 79)

- (7) *La-f-o-töi danö hö’ö taro’o zi’o.*
 3PL.RL-CAUS-HAVE-name land.MUT DIST plant stick.MUT
 ‘They called that land *taro’o zi’o*.’

Korowai (Awyu-Dumut; van Enk and de Vries 1997: 90)

- (8) a. *ate-lé*
 take-1SG
 ‘I take/took’
 b. *ate-lè.*
 take-1PL
 ‘We take/took’

Rigid word order was defined so that it occurred when a change in the order of arguments altered the thematic interpretation of the sentence (Primus 1999: 132). For instance, in the English sentence *The mule kicked the horse*, *the mule* functions as A and *the horse* as P, but this sentence has a different thematic interpretation from the sentence *The horse kicked the mule*, where *the horse* functions as A and *the mule* as P. Rigid word order may also be restricted to a particular context: in Norwegian (9), sentences with two animate arguments have rigid word order, but non-rigid word order otherwise.

Norwegian (Germanic; Øvrelid 2004: 1, 7)

- (9) a. *Brevet skrev jenta.*
 letter.DEF write.PST girl.DEF
 ‘The letter, the girl wrote.’
- b. *En nabo oppdaget innbruddstyven.*
 a neighbor discover.PST burglar.DEF
 ‘A neighbor discovered the burglar.’
 (*The burglar discovered a neighbor.)

Three restrictions were further observed in defining the limits of core argument marking. First, the coding strategies were researched in simple active, affirmative, declarative main clauses, where active means the most neutral clausal configuration, excluding passives, for instance. Second, the coding strategies were examined in the clause proper, excluding sentences in which a transposed element was separated from the rest of the clause by a pause or marked by a pronoun in situ (cf. Siewierska 1998). Third, the focus was mostly on noun arguments (Articles 1 and 3), because the high discourse frequency of pronouns increases the conservativeness of their word order and morphological properties in comparison to nouns (Bybee and Thompson 1997).

The domain of inquiry varies slightly among the articles. Article 1 focuses on the interaction between the coding strategies, Article 2 on the functional domain as a whole, and Article 3 on the interaction between morphological marking and word order.

2.5. Complexity in the case studies

In this section I describe how I measured complexity in the case studies (Sections 2.5.1-2.5.3) and how I evaluated the importance of different complexity types (Section 2.5.4).

2.5.1. Article 1: Complexity trade-offs in core argument marking

In Article 1, I researched whether there were any complexity trade-offs between head marking, dependent marking, and rigid word order in core argument marking.

Complexity was then defined as the degree to which the particular coding strategy discriminated the core arguments from one another. This definition corresponds to the functional load of a coding strategy (see Surendran and Niyogi 2006: 43). Functional load is related to complexity in that the more fully a certain coding strategy differentiates the arguments, the more redundant is the strategy, because animate and/or definite A as well as inanimate and/or indefinite P can typically be identified by virtue of their semantic properties alone (cf. Croft 2003: Chapter 6.3; Haspelmath 2008: 13-14; Sinnemäki, forthcoming). As a result, from the viewpoint of the case system, there is a correspondence between the great functional load of a coding strategy and the high degree of redundancy and thus of system complexity (in terms of Dahl 2004). In addition to functional load, complexity was also measured as overt coding (the presence vs. absence of a coding strategy). As suggested in the Commentary to Article 1, I now consider this a more analytical metric and preferred over functional load. In Sinnemäki (2005, 2006), I also measured the conceptual complexity of the coding strategies, but as that did not produce any significant results, I have not pursued the measuring of conceptual complexity here (see Sinnemäki, forthcoming, for a way of adapting conceptual complexity to the differential case marking of the object).

Throughout the research for this dissertation, a recurring question has been why I consider rigid word order to be of greater complexity than non-rigid word order and not the other way around. The argument behind this objection is as follows: rigid word order is more predictable and less informative than non-rigid word order (or free word order); therefore, would not rigid word order then be of lower complexity than non-rigid word order? The question is pertinent, given that Fenk-Oczlon and Fenk (2008: 60-63) themselves disagree on this question. Because I have not thoroughly addressed this issue in the articles, I will simply outline below the basic reasons why I think rigid word order is more complex than non-rigid word order in argument discrimination.

First, the answer crucially depends on the particular approach to complexity. If the focus is on measuring the entire information content in an object, then the object with the greatest amount of variation (or randomness) is the most complex (cf. Footnote 9). For word order, this means that its complexity would be equated with the degree of word order variation, and the system with the greatest amount of variation would be

deemed the most complex in that regard. The problem with this interpretation lies in the nature of that variation, namely, whether it is specified for a particular functional purpose or not. If it is not, then that word order variation is nothing but meaningless random variation, which essentially has no structure and thus very low effective complexity. From the viewpoint of effective complexity, a system with an unconstrained concatenation of elements is minimally complex, and any constraints imposed on the concatenation of elements increases complexity (Dahl 2004: 52). In short, what increases complexity in my approach is the number of constraints imposed on word order.

However, the emphasis on effective complexity is insufficient in itself to explain why I consider rigid word order more complex than non-rigid word order. The reason is that there is probably no such thing as unlimited concatenation in languages, because even the most seemingly unconstrained word order tends to be employed for pragmatic purposes (e.g., Payne 1992). In that sense, word order variation is often, if not always, specified for a certain purpose in languages, and, therefore, a language with great word order variation would have greater effective complexity of word order than a language with more invariant word order. It is here that my suggestion of focusing on particular types of complexity in their local contexts helps to clarify the issue.

My second reason for considering rigid word order more complex than non-rigid word order has to do with my emphasis on measuring different types of (effective) complexity in their local contexts rather than measuring overall complexity (see Section 2.3.2). In the case of word order, this means measuring complexity separately in the functional domains where it is used, for instance, in core argument marking, in marking information functions (Payne 1992), or in the parts-of-speech systems (Hengeveld, Rijkhoff, and Siewierska 2004). Because my focus is on the role of word order in core argument marking, my definition for rigid word order is restricted to this domain only and is not meant as an overall measure of word order (in)variance. In Section 2.4, I defined rigid word order in relation to the thematic interpretation of the sentence so that the unconstrained order of arguments is a sign that word order has no role in core argument marking, whereas the presence of such constraints indicates that word order has a role in that domain. As discussed above, the presence of constraints on argument

order increases the description length of word order as a coding strategy in the domain of core argument marking, so from this perspective, rigid word order has greater (effective) complexity than non-rigid word order. Combining effective complexity with a focus on different types of complexity in their local contexts thus helps to clarify the contradictory views on the complexity of word order.

2.5.2. Article 2: Complexity in core argument marking and population size

In this article, I studied whether complexity in core argument marking correlated with population size. Complexity was defined in terms of adherence to vs. departure from the principle of one-meaning–one-form in core argument marking. A language was classified as adhering to this principle if the arguments were discriminated from one another in roughly all contexts via the three coding strategies of head marking, dependent marking, and/or rigid word order. A language was classified as violating this principle if the arguments were marked with multiple coding strategies (violation of economy) or if no coding strategies were used in some contexts (violation of distinctiveness). For instance, Maybrat used rigid word order in all contexts; therefore, Maybrat was classified as adhering to the principle of one-meaning–one-form (note that the head marking was counted only if both arguments were head-marked, thus discarding the head marking of one argument). West Greenlandic, for its part, used both head marking and dependent marking, but also in some contexts rigid word order, so West Greenlandic was classified as violating economy, while Iau used dependent marking only in restricted contexts, so it was classified as violating distinctiveness (personal communication with Janet Bateman).

The principle of one-meaning–one-form is related to complexity in that adherence to this principle requires a shorter description than its violation (Miestamo 2008; see also Section 2.3.6). The relationship between form and meaning is one-to-one when adhering to this principle. It thus exhibits low organizational complexity between the mapping of meaning and form (or low system complexity in terms of Dahl 2004). Violation of economy is a matter of a many-to-one relationship between form and meaning, while violation of distinctiveness is a matter of a one-to-many relationship.

Organizational complexity between the mapping of form and meaning thus increases when the principle of one-meaning–one-form is violated, as the relationship between form and meaning has to be specified in more detail. Observe, however, that my metric does not take into account the different degrees of violation.

2.5.3. Article 3: Word order in zero-marking languages

In Article 3, I researched whether the absence of the morphological marking of core arguments, that is, zero marking, correlated with SVO word order. Zero marking is clearly less complex than overt marking in terms of compositional complexity: it has lower constitutional complexity, since no morphemes are coding the arguments, but it also has lower taxonomic complexity, since the whole morphological category (e.g., of case) is absent from the language. In terms of Dahl (2004), zero marking has lower structural and system complexity than overt marking.

When it comes to SVO word order, I argued that the preference for zero marking occurring in SVO languages has to do with SVO being a more economical and iconic order than any other ordering of S, O, and V. I proposed that verb-medial orders have more economical argument linking than **verb-terminal** (verb-initial and verb-final) orders and that to position S before O is more iconic than to position O before S. However, throughout my research, there has been the recurring question of why I consider SVO (or verb-medial order) less complex than the other orders from the viewpoint of information theory and whether the issue has anything at all to do with complexity. Since I have not explained my position clearly enough in Article 3, I explain it here in detail.

Because my focus in this dissertation is on the complexity of coding strategies in argument marking, it is necessary to show whether and in what way the different orders of S, O, and V may differ from one another in terms of complexity. My main reason for considering one order more complex than another is that the description requires more constraints to be used (cf. Dahl 2004: 52). As discussed above, unconstrained concatenation is minimally complex, and the addition of constraints on the order of elements increases description length. When comparing the different orders of S, O, and

V to one another, the idea is to consider separately the orders between S and V, between O and V, and between S and O. To study clausal word order with these parameters is justified on empirical grounds as well (Dryer 1997). When the focus is on these three pairs of orders, the complexity of different word orders appears as follows.

Regardless of which order of S, O, and V a language uses for argument discrimination, the order of S and V and that of O and V must be constrained. The reason is that the verb serves as the reference point for linear order, so the order of arguments must be constrained in reference to the verb (Frajzyngier and Shay 2003: 58-59). This means that every order of S, O, and V is limited by at least two constraints, and the difference between the orders is a matter of additional constraints. If a language has SV and OV orders, this yields SOV and OSV orders, and if a language uses VS and VO orders, then these yield VSO and VOS orders. An additional constraint is thus needed, for instance, in languages with SOV or VSO order, so that the word order can be a viable means in core argument marking. That constraint has to do with the order of S and O (it is this constraint which further prevents the use of pro drop in zero-marking verb-terminal languages, for example; cf. Dryer 2002). However, in SVO languages the orders SV and VO already specify the relative order of S and O, so it is unnecessary to specify their order separately (the same goes for OVS languages). This results in a shorter description and thus in smaller effective complexity of rigid word order in verb-medial languages, because it requires two constraints compared to three in verb-terminal languages. The simplicity of SVO (and OVS) order is therefore a matter of a smaller number of constraints on argument order. This is essentially what is meant in SVO languages when it is said that the arguments can be identified based on their position relative to the verb alone (cf. Dryer 2002; Frajzyngier and Shay 2003; Hawkins 2004).

One further point needs to be added: strictly speaking, it is incorrect to say that SVO is the simplest word order available for argument discrimination, because OVS is equally simple in information-theoretical terms. The correct generalization is that verb-medial word order is simpler or more economical than verb-terminal word order. I argued for this generalization already in Article 3, and it served as a basis for bringing in the iconicity of sequence as a necessary additional factor for explaining the results. S-O is thus a more iconic word order than O-S, because it follows the order initiator-target

in the real world and is therefore also less costly to process. All in all, while SVO is not the simplest word order in core argument marking, it is the most optimal, as it is supported by both economical and iconic motivations, not just by one of them.

2.5.4. Evaluation

The most important types of complexity in the case studies were structural complexity, system complexity, and conceptual complexity in terms of Dahl (2004), and compositional, taxonomic, and organizational complexity in terms of Rescher (1998). Dahl's notions of system complexity and conceptual complexity played a major role in Article 2, since both can be used to measure different aspects of deviation from the principle of one-meaning–one-form. Deviations from the one-to-one mapping between meaning and form increase system complexity, because the mapping is no longer fully transparent, but these deviations also increase conceptual complexity, because the conditions for overt marking must then be specified. A similar definitional correlation also occurred between system complexity and structural complexity, because both can be used to measure the presence vs. the absence of overt coding (Section 2.3.3).

As for structural complexity, it played a major role in Articles 1 and 3, but a minor role in Article 2. In this sense, there was a greater focus in my case studies on structural complexity than on conceptual complexity. Hierarchical complexity was virtually untouched, owing to its small role in core argument marking (see Givón 2009 for a recent functional-typological treatment on its development in languages). The relevance of these notions of complexity to language universals is discussed in Section 4.3.3.

Chapter 3 Sampling, statistical methods, and data

In this chapter, I outline the sampling methods and statistical tests that were used in the case studies (Section 3.1) and describe the data (Section 3.2).

3.1. Evaluation of language universals

3.1.1. On sampling

Typologists cannot talk about language universals if the generalization at hand cannot be extrapolated or generalized beyond the sample, whether to all currently spoken languages or to all possible languages. To be able to generalize beyond one's data, the data have to be collected, that is, sampled, in a proper way. The issue of language sampling is essential to the quest for language universals, and it has raised a great deal of discussion during the previous thirty years (e.g., Bell 1978; Dryer 1989, 1992, 2000; Perkins 1989; Rijkhoff et al. 1993; Rijkhoff and Bakker 1998; Maslova 2000; Bickel 2008a, 2008b, 2011). Greenberg (1966) famously did not use random or stratified sampling, and the size of his sample was quite small by modern standards ("only" thirty languages, but this did not preclude robust results in his case; cf. Haspelmath and Siegmund 2006). Generally, though, a convenience sample of this kind is the worst possible sample for making reliable generalizations (see Widmann and Bakker 2006).

When the data seem to have revealed an interesting distributional pattern, they may have been affected by several factors. Sir Francis Galton (1822-1911) first noted that similar behavior between variables may be caused by three basic factors: borrowing (areal diffusion), inheritance (genealogical relatedness), and universal effects. What has come to be called **Galton's Problem** is that, if one does not control for the possible confounding effects of borrowing and inheritance (and perhaps other factors as well), then the observations are not independent of one another, and the results do not necessarily show a genuine universal trend. This problem of confounding factors, also called autocorrelation, is common to all disciplines.

The classical way of controlling for the confounding factors is to devise a stratified random sample, which then enables making statistical inferences from the

sample to the population (e.g., Perkins 1989). However, random sampling is rather limited and difficult in typology: on the one hand, sampling at the lowest taxonomic levels may inflate areal effects, while, on the other hand, stratified random sampling is impossible in the case of language isolates (Bickel 2008a: 222; Janssen, Bickel, and Zúñiga 2006: 420-424). Random sampling is thus not an ideal option in typology. When it is used, it is important to note that it makes inference to the currently spoken languages only, providing static snapshots, so to speak, of the variable's distribution. I call such static snapshots **cross-linguistic trends** or **synchronic universals**.

Stratified random sampling was used to compile the samples in Articles 1 and 2. This means that the data in these articles warrant inferences to currently spoken languages, but not to all possible human languages. These samples were stratified both genealogically and areally by sampling a number of genera from each of the six macro-areas of Dryer (1992) at random and in proportion to the genealogical diversity in that area (see Miestamo 2005: 31-39). The genealogical effects were controlled by sampling maximally one language per genus, which controls for the most obvious genealogical bias, since closely related languages are not allowed in the same sample. This type of sampling is here called **controlled genealogical sampling**, or just g-sampling (Bickel 2008a). A further control for non-independence was done by modeling the data with multiple logistic regression and factoring areality in the models themselves (in Articles 1 and 2 this was done in the commentary on these articles, because I was not acquainted with multiple logistic regression at the time of writing).

In Article 3, I used a different strategy for controlling the confounding factors. A growing trend in typology is to conceptualize language universals as structural pressure on how languages change over time; in other words, as diachronic laws of type preference (e.g., Greenberg 1978, 1995; Maslova 2000; Croft 2003; Maslova and Nikitina 2007; Bickel 2008b, 2011). This view acknowledges that there is a further dimension to be controlled for, namely, the possible dependence of synchronic typological distributions on distributions in earlier language populations (see especially Maslova 2000). Distributions at different times are considered to be different states, and changes between the states are modeled by estimating the probability for a particular type in the typology to change to another type. If this **transition probability** is rather

low, then there would be a detectable dependence between the distributions at different states, meaning that the synchronic typological distribution at hand could reflect the effects of inheritance over deep time and not genuine universal effects.

For the purpose of assessing whether there is universal diachronic pressure for languages to change in a certain direction over time, it is necessary to sample languages both within and across families (family is here understood as any genealogical grouping of languages in the taxonomy). This method thus takes a very different approach to sampling compared to (stratified) random sampling and treats each language as a potential independent case of retaining or changing a feature of the proto-language (Maslova and Nikitina 2007). This also means that using large samples does not pose similar problems for statistical modeling, especially when the confounding factors are modeled with multifactorial statistical tests (see below).

Transition probabilities can be estimated for a given typology by examining variation within language families. This can be done by counting the number of **type-shifts** in the language population, that is, by counting the number of languages changing towards a particular type and the number of languages changing away from that type (Maslova 2000; Maslova and Nikitina 2007). Based on these probability estimates, it is then possible to calculate an ultimate stable state where the transitions to a type and away from that type cancel each other out. In that stable state, language change no longer alters synchronic frequencies, since they are in equilibrium, determined by the ratio of the transition probabilities. Those probability estimates may then ultimately reflect the effect of different forces that have shaped the synchronic distributions. I have not used this method here, mainly because it provides no way of estimating the impact of language contacts on typological distributions (see also Bisang 2004a).

An alternative method for approaching transition probabilities is the **Family Bias Theory** of Bickel (2008b, 2011) (called the Family Skewing Method in Bickel 2008b). In this approach, transition probabilities are not directly assessed, but their effect is estimated by observing the impact of family internal variation on typological distributions. If that impact is small enough, then it can be removed, and we can infer that the resulting distribution is independent of family internal variation within the observable time-frame. This approach lends itself naturally to multifactorial statistical

modeling, and it also enables an assessment of the effect of language contact or social factors separately. These approaches are briefly discussed in Article 3, but a fuller comparison of their pros and cons would require an altogether separate study.

There is a potential problem, however, in these two approaches to transition probabilities, namely, that these methods may be quite sensitive to birth-and-death processes within families, even though they may not be sensitive to these processes in large language populations (cf. Maslova 2000; Maslova and Nikitina 2007). A similar problem might be caused by sampling decisions within families. To evaluate the extent to which these sensitivities constitute real problems for estimating transition probabilities would require a separate study, one that might assess these effects via computer simulations. This was not possible within the limits of this study and must be left for future research.

This dynamic way of modeling language universals enables (logical) inferences to be made from the sample to all possible languages to the extent that the sample is representative and to the extent that the typological distribution is significant and independent of confounding factors. Distributions that meet these criteria are here called **dynamic universals**, a term I use to refer to the independence of typological distributions from known diachronic influences within the timeframe provided by the comparative method. The effect of the confounding factors was estimated by modeling their effect via multiple logistic regressions, discussed in the following section.

As mentioned above, I use two approaches for controlling the confounding effect of genealogical relatedness, the g-sampling method and the Family Bias Theory. In Articles 1 and 2, I used only g-sampling, but in Article 3, I modeled the data with both methods to contrast a basically synchronic approach to universals (g-sampling) with an approach that provides a dynamic interpretation for universals (the Family Bias Theory). There are at least three reasons for doing so. First, the Family Bias Theory scrutinizes statistical trends within families, and, thus, **singleton families**, that is, language isolates and single sampled members of a family, are omitted from the outset. To examine the distributions with g-sampling enables these families to be included.

Second, there is a growing tendency in linguistics to approach research questions **multi-methodologically**, especially when there is no consensus as to the best method or

the right way of variable encoding (e.g., Arppe and Järvi­kivi 2007; Bickel 2008b; Kertész and Rákosi 2008; Haspelmath 2009; Arppe et al. 2010). For this reason, to contrast two different ways of modeling genealogical relatedness captures variation in typological distributions in a more comprehensive way than by adhering to only one approach. In some instances, this contrast may also reveal places in which synchronic variation departs from an otherwise dynamic universal. Bickel (2011) suggests an alternative way to include singleton families, namely, via extrapolation from known to unknown families by assuming that “[u]nknown families are subject to the same developmental principles as known families” (p. 12). While this is a promising way to include singleton families under dynamic universals, the possible limitations of this “Uniform Development Assumption” are poorly known. Detailed computer simulations in future research might provide a way to assess the limits of this assumption.

Third, to use different ways of coding the variables—the types in the typology, areas, or families, for instance—is a simple step towards reproducibility in typology, whose importance has recently been strongly emphasized (e.g., Bisang 2004b; Corbett 2005: 19-20; Harris, Hyman, and Staros 2006; Song 2006: 8-15). It is also a step toward making the results less dependent on one method or sampling technique. Newmeyer (1998: 297ff), for one, doubts whether the results of functional-typological research are mere artifacts of a less than optimal sampling method. If results are independent of sampling choices, this helps to dispel such doubts, at least to some extent.

These three issues increase the importance of methodological issues in typology, but it may come at a cost, because emphasizing the role of methodology may prevent the testing of multiple hypotheses in a single study. Yet research often progresses from an initial modeling of hypotheses with limited data to extending the model to larger samples, and finally to replicating the results with new data and alternative methods. To increase the robustness of the results in linguistics, multi-methodological approaches are thus called for instead of methodological monism (*pace* e.g., Sampson 2007).

3.1.2. Statistical methods

The statistical methods used in my work can be classified in **monofactorial methods** and **multifactorial methods** in terms of Gries (2003). The former methods include the binomial exact test, the chi-square test for independence, Fisher's exact test, and the ordinal correlation tests, namely, Kendall's tau and Somers' d. One multifactorial method—multiple logistic regression—was also used. Table 3 shows the methods used in each case study. Monofactorial tests are useful for testing the effect of individual variables in isolation, but because linguistic phenomena are affected by numerous coincident variables, their modeling naturally calls for multifactorial tests that estimate the effect of independent variables in competition (Gries 2003: 79). The details of these methods are not discussed here, since they are properly described in my articles.

The monofactorial tests were used for assessing such things as whether the variables of interest were significantly associated with each other. The multifactorial methods were used for modeling the effect of multiple variables simultaneously, estimating whether a given correlation was independent of confounding factors, and determining which main effects had a significant impact on the variable of interest. A crucial component in statistical tests is a test for significance. Owing to the limitation of many statistical tests, such as chi-square, the significance of the results was in most instances deduced by means of a randomized Monte Carlo permutation test, following recommendations by Janssen, Bickel, and Zúñiga (2006). Computation and plotting were done in the open-source statistical environment R (R Development Core Team 2010 or earlier versions), using additional packages *Design* (Harrell 2010), *Hmisc* (Harrell et al. 2010), and *vcd* (Meyer et al. 2006, 2010).

Case study	Statistical tests used
Article 1	Chi-square test for independence; Kendall's tau (logistic regression in the commentary).
Article 2	Chi-square test for independence; Fisher's exact test (also Somers' d and logistic regression in the commentary).
Article 3	Binomial exact test; Fisher's exact test; logistic regression.

Table 3. Statistical methods used in the dissertation.

In typology, as in social sciences in general, the results of statistical tests are usually considered significant if the probability estimate is smaller than the critical value 0.05. While such limits are important for estimating the meaningfulness of the results, I would not attach too much importance to them, but rather emphasize the tolerance of the results of issues of sampling and misclassifications. In typology, the sample sizes are often small, so the p-values that come close to the critical value 0.05 may be rather vulnerable to such things as misclassifications. In that sense, it does not matter very much if the method used has high or low power, because if the p-value is only slightly less than the critical value, the result often becomes non-significant, even if a single data point is altered. For this reason, I used several **reliability tests** in the spirit of Janssen, Bickel, and Zúñiga (2006: 435-438) to estimate the reliability of the results in each article. In Article 1, I used bootstrap resampling; in Article 2, a reliability test adapted from Janssen, Bickel, and Zúñiga (2006); and in Article 3, various sensitivity analyses and alternative methods for modeling the data. These tests confirmed the results and showed that the main results were robust against misclassifications or issues of sampling.

3.2. Data

In empirical studies of language universals, the data are typically drawn from reference grammars and works dealing with the particular variable of interest. Field work, if done, is generally limited to a few languages (a few dozen at most) for the practical reason of the sheer quantity of work that goes into data gathering. Thanks to the recent publication of typological databases (e.g., the *WALS*; Haspelmath et al. 2005) and genealogical taxonomies (e.g., *Autotyp* genealogical classification; Nichols and Bickel 2009) and the growing trend of online publication, typological work of this kind has become much easier just in the last few years by virtue of better data availability. The present dissertation has also benefited enormously from these new developments.

The most important sources in my work were grammar descriptions (e.g., Dol 1999 on Maybrat) and other scholarly texts on core argument marking. Publicly available databases, such as Nichols (1992) and Comrie (2005), were used in Article 3

to analyze the morphological type of a given language, but these sources were used sparingly. In a few instances, the data were obtained through personal communication with language experts. However, most of the data on dominant word order in Article 3 were taken directly from Dryer (2005b).

As for the sample composition, the data for Article 1 came from an areally and genealogically stratified random sample of fifty languages, while those for Article 2 came from a slightly modified sample of Article 1 (see the Commentary on Article 2). The data for Article 3 came from a sample of 848 languages, covering roughly 94 percent of genera in the *WALS* genealogical taxonomy (Dryer 2005a).

The reliability of the data is always an important question in typology, since secondary sources play a big role. In my study, the analysis of overt marking patterns can be considered rather reliable, since information on case marking and agreement is easily available in the sources (but see the Commentary on Article 1). Data on rigid word order were more difficult to obtain, even from otherwise reliable reference grammars. This may pose a slight problem for the results of Articles 1 and 2. This problem was discussed in part in the Commentary on Articles 1 and 2, where I reanalyzed some of the data based on sources of which I was unaware at the time of writing those articles. These changes do not constitute a threat to the results; on the contrary, they strengthen them. A possible way around the problem of data availability on rigid word order would have been to devise a typological questionnaire, as in Siewierska (ed., 1998). However, within the practical limits of this dissertation such a solution did not seem to be a feasible option.¹⁴

On the whole, the data analyses and coding in this dissertation were done in the spirit of the *WALS* (Haspelmath et al. 2005). Recently, Dixon (2009: 257-263) has criticized the use of poorer than necessary sources for enterprises such as the *WALS*, but a simple, practical reason may explain the apparently selective use of sources, at least in part, namely, that the sources are unevenly distributed in research institutions and libraries around the world. It may be financially prohibitive to purchase copies of all relevant works, or to travel to research institutions where a great number of sources is

14 While questionnaires may at best provide more reliable data than reference grammars, they present their own issues of unreliability (see Croft 2003: 29).

available (e.g., the Max Planck Institute for Evolutionary Anthropology in Leipzig). In the future, a simple solution to this problem is the digitizing of especially hard-to-obtain reference grammars, made available for free or by subscription.

There is a growing trend in typology to base language comparisons on parallel texts, such as translations of the Bible (Cysouw and Wälchli 2007; Wälchli 2007). Analysis of parallel or original texts has been applied in complexity research as well (Juola 1998, 2008; Bane 2008; Moscoso del Prado Martín 2010, 2011). This so-called **original-text typology** is a new and interesting approach for future typological research, because, when used as a complement to grammar-based typology, a more complete picture of grammatical variation can be achieved both within and across languages. For practical reasons, it was not possible here to complement my grammar-based analysis with the methods of original-text typology, but that would be an interesting angle for future studies of language complexity.

Chapter 4 Results and discussion

This chapter summarizes the main results of my dissertation, classified as three types: i) methods for measuring language complexity in a cross-linguistic way (Section 4.1), ii) methods for evaluating language universals (Section 4.2), and iii) the empirical results (Section 4.3). Explanations of the empirical results are discussed in Section 4.4.

4.1. Complexity in typology

One of the main methodological contributions of my dissertation is that the approach to language complexity proved useful when applied to empirical data. This approach was comprised of the following aspects: i) complexity is separated from difficulty; ii) complexity is measured as the description length of an object's structure, that is, as effective complexity; iii) the emphasis is on local complexity rather than global complexity; iv) the general notion of complexity is broken up into different types (or modes), in the spirit of Rescher (1998); and v) complexity is most fruitfully measured when the complexity differences are rather large (following Gell-Mann and Lloyd 2004 and Miestamo 2008: 25), which enables the use of tools developed by linguists rather than by mathematicians or computer scientists (e.g., Vulcanovic 2007). This approach takes into account the limits imposed on language complexity research by the typological approach. In effect, the crux of my approach is to measure particular types of complexity in their local contexts (e.g., in the same functional domain) between variables that clearly differ from one another in terms of complexity.

An important part of understanding complexity is how to translate the analytical framework of one approach to the vocabulary of complexity (Hübler 2007). I have argued, following especially Kusters (2003) and Miestamo (2008), that principles common to functional-typological research, such as economy, distinctiveness, and one-meaning-one-form (or transparency), can be fruitfully related to complexity (Section 2.3.6). The use of these principles as complexity metrics (as in Article 2) is thus not a forced idea, but the result of understanding how the principles relate to complexity. In

this sense, our understanding of these common principles has also increased as a consequence.

A crucial question for language complexity research is the extent to which the notion of complexity is relevant for the field at large. Gil (2008: 129-130), for one, is doubtful as to the relevance to typology, seeing this notion as being more relevant to phylogeny, sociolinguistics, diachrony, and language acquisition research. It is true that these subfields have applied the notion of complexity in fruitful ways (see Bichakjian 1999; Longa 2001; Locke 2008; Dahl 2004; Trudgill 2004; Nieminen 2007; Ellis and Larsen-Freeman 2009). However, my results suggest that complexity is relevant for typology as well, especially in that the interrelationships among linguistic variables can be found in terms of complexity and that complexity can increase our understanding of key principles in the field. In addition, since cross-linguistic comparison requires clear definitions, the development of such definitions bears fruit, on the one hand, in increasing our understanding of language complexity and, on the other hand, in benefiting those subfields which apply the notion of complexity more than typology does at present.

In Section 2.1 typology was characterized as a subdiscipline that studies cross-linguistic diversity and how linguistic patterns interact among themselves as well as vis-à-vis other anthropological patterns. Given this characterization, there may be even greater relevance for the notion of complexity in typology than suggested above. First, there has recently been a growing interest in the interaction between complexity and diversity. While these notions represent different things, they are related; on the one hand, diversity seems to be required for complexity (Page 2011), and on the other hand, complexity seems to generate diversity (Doebeli and Ispolatov 2010; see also Rescher 1998). These issues are suggestive of a relationship between cross-linguistic diversity and language complexity as well.

Second, one of the most important ingredients of complexity is the interaction between patterns in a system (Page 2011; Section 2.3.3). This means that in determining how linguistic patterns interact among themselves or with other anthropological patterns, we are essentially contributing to our understanding of language as a system that has many parts, which interact and adapt or, in other words, to viewing language as

a **complex (adaptive) system** (cf. Ellis and Larsen-Freeman 2009). These issues mean that research into cross-linguistic complexity can contribute in a meaningful way to the core agenda of modern typology.

4.2. Evaluation of language universals

Multifactorial methods have not often been used in typology for testing typological distributions. Methods such as multidimensional scaling and cluster analysis have been used to assess the clustering of typological variables or the genealogical relatedness of languages (e.g., Cysouw 2007; Dunn et al. 2005, 2008), but they have been used less often in testing language universals. Multiple logistic regression has rarely been used in typological work, with notable recent exceptions by Balthasar Bickel and his coworkers (e.g., Bickel 2008b; Bickel and Witzlack-Makarevich 2008; Bickel, Hildebrandt, and Schiering 2009), by Cysouw (2010), and by Bakker et al. (2011). The introduction of logistic regression to the study of language universals in my dissertation may serve as an initial prototype and a springboard for the further application of logistic regression in typology.

I also adopted the insight of universals as diachronic laws of type preference, as implemented in Bickel (2008b, 2011). This approach enables modeling language universals in a way that provides dynamic interpretations for the results, even extending the generalizations to all possible human languages. It is a matter for future research to determine the limits of this approach, but currently, this may be the most promising way to examine dynamic patterns in synchronic typological distributions.

Besides this dynamic approach, I implemented the g-sampling for modeling synchronic cross-linguistic trends. Because there is no consensus in the field as to how genealogical relatedness should be controlled for, I opted for this kind of a multi-methodological approach, contrasting the two ways of modeling genealogical relatedness. I extended this approach to the modeling of geographical areas as well, by contrasting several different areal breakdowns at different levels of granularity (Article 3). The main reason for such contrasting was that often we do not know at which level of granularity the effect of areal diffusion could be expected in each instance, the

possibilities varying anywhere between two large macro-areas (e.g., the Circum-Pacific vs. the Old World in Bickel and Nichols 2006) and at least two dozen small areas (e.g., in Bickel 2008b). This multi-methodological approach could be extended to the coding of structural types as well. Koptjevskaja-Tamm (2010), for one, entertains the idea that “[l]inguistic phenomena normally lend themselves to multiple classifications” (p. 585), meaning that one and the same typological variable can be classified in multiple ways, providing sometimes diverging results, sometimes converging results.

There are a few more benefits to contrasting the g-sampling and the Family Bias Theory when testing typological distributions. First, to the degree that the results produced by these two methods converge, the conclusion is stronger than using either of these methods alone. And if the results produced by these methods tended to converge in the great majority of cases, this would suggest that g-sampling would quite accurately approximate the results, or at least the conclusions, produced by the Family Bias Theory.

Second, other benefits occur when the results produced by these methods diverge. If, for instance, the results produced by g-sampling showed local paths of grammaticalization, but those produced by the Family Bias Theory showed universal paths, this could mean that g-sampling was able to capture some aspect of instability in the variable’s distribution. If, however, the results diverged in the other way, then it could indicate, for example, that singleton families behave differently from non-singletons, which would call for more detailed studies of the variable’s distribution.

I also took seriously the call by Janssen, Bickel, and Zúñiga (2006) to implement permutation methods and sensitivity tests in typology. Using permutation methods or sensitivity tests is no substitute for accurate data analysis, but the latter type of tests especially provide at least some estimates of the limits of the results’ reliability and, in this sense, are superior merely to noting some vague possible problems in the data.

4.3. Empirical results

The empirical results of my dissertation are discussed next in light of the general research questions stated in Chapter 1 and reiterated below.

- Is there any systematic cross-linguistic variation in the grammatical complexity of languages in a particular domain?
- Can grammatical complexity be affected by the social environment of a speech community, for instance, by population size?

These questions were divided into three, more detailed questions in the case studies. The results related to systematic complexity variation in core argument marking are dealt with in Section 4.3.1, and the results related to complexity variation dependent on geographical or social environment are dealt with in Section 4.3.2. Section 4.3.3 briefly discusses the relevance of different notions of complexity to language universals.

4.3.1. Question 1: Systematic complexity variation

The question of whether complexity varies in a systematic way in core argument marking was investigated in Articles 1 and 3, discussed here in that order. Both articles take up correlations touted as being universal in the scholarly literature, but these correlations have not been tested with large and balanced samples.

ARTICLE 1: COMPLEXITY TRADE-OFFS IN CORE ARGUMENT MARKING. According to the results of this article, complexity trade-offs occurred among the coding strategies that interact in core argument marking, but only in a limited way. A negative correlation existed among all three coding strategies, but only the correlation between dependent marking and rigid order was significant, while head marking did not correlate with rigid order or dependent marking in a significant way, regardless of how complexity was measured (as functional load or as overt vs. absent marking; see the Commentary on Article 1). These results suggest that there is a close relationship between dependent marking and rigid order, but none between head marking and dependent marking, on the one hand, or between head marking and rigid order, on the other. Head marking seemed to have a “backup function” in that, if a language did not distinguish the arguments in all contexts, then head marking (at least of P) was present; its role in argument

discrimination thus seems to be marginal and largely redundant from a typological point of view.¹⁵

It has often been noted that there is a close relationship between rigid order and dependent marking (e.g., Sapir 1921; Vennemann 1973, 1974; Lehmann 1978; Kiparsky 1997; Blake 2001: 158). However, this relationship has not been researched typologically and so has remained a largely unverified hypothesis. In Article 1, this hypothesis was tested and confirmed with typological data.¹⁶

In addition, while there was a significant relationship between rigid word order and dependent marking, the presence of some counterexamples falsifies their relationship as an exceptionless universal (cf. Pinker and Bloom 1990). This is not detrimental to a functional-typological approach to grammar, where it is assumed from the outset that grammatical structure can be affected by numerous forces that exert their effect in a probabilistic way. However, such counterexamples call for reassessment of assumptions that claim an exceptionless universal status for certain correlations or grammatical features, as recently emphasized by Evans and Levinson (2009).

What was rather unexpected in the results was that head marking correlated neither with rigid order nor with dependent marking. According to Siewierska (1998), the degree of word order variation depends more strongly on agreement than on case marking, so one would have expected head marking to correlate with either variable in my research as well. After the publication of Article 1, I checked whether head marking correlated with either variable if defined differently, that is, if person marking on the verb had been excluded from its definition, since person marking alone cannot discriminate between the arguments if both are in third person. However, preliminary results suggested that this did not strengthen the correlations and that the original results were not a matter of definition. My tentative explanation for the absence of correlation

15 However, in Sinnemäki (in preparation) I show that head marking correlates negatively and significantly with dependent marking of person indices, that is, pronouns or person marking on the verb (data from the *WALS*; Haspelmath et al. 2005). It thus appears that the absence of any correlation between head marking and dependent marking is limited to noun arguments.

16 Siewierska (1998) studied the dependence of word order variation on case marking or agreement, but not the use of rigid order for argument discrimination, which is a slightly different matter.

between head marking and the two other coding strategies is that, while the latter can usually link the thematic structure to particular NPs quite efficiently, head marking by itself can do the job only in a limited way, depending on the degree to which it can unambiguously identify the thematic role of a particular NP argument. Its efficiency often depends on the use of rigid order or dependent marking, and thus it is unsurprising that there was no correlation (but see Footnote 15).

Another important, although less central, result of Article 1 was the observation that languages allow violation of economy more than violation of distinctiveness. This result was corroborated by the reanalyzed data for Article 2, which indicated that more than half of the languages violated economy (see the Commentary on Article 2). In other words, redundant marking is tolerated more often than the absence of distinctive marking in core argument marking. This result agrees with the observation that languages tend to retain complex grammatical marking and distinctions that appear redundant for efficient communication (cf. McWhorter 2001, 2007; Locke 2008).

Second, the fact that economy was violated more than distinctiveness suggests that the role of economy as a principle in shaping linguistic structure or representations may require some slight readjustment. Economy, or minimality, is crucial in linguistic theorizing, largely regardless of the researcher's goals or theoretical framework (e.g., Zipf 1948; Haiman 1983; Chomsky 1995; Deutscher 2006; Haspelmath 2008). Yet when it comes to language processing, minimality does not seem to be a basic principle, but rather a special case of a more general principle of distinctiveness (Bornkessel-Schlesewsky and Schlewsky 2009). Consequently, economy should be more clearly defined than before in relation to distinctiveness, along the lines of "minimize what you can" instead of "minimize all" (*pace* e.g., Fodor 1998). Emphasizing the role of economy too much may also conceal the utility of some type of redundancy for communication in providing important cues for the hearer as well as in increasing the robustness of grammar (e.g., Keller 1994: 108-114; Kusters 2003: 46-52; Dahl 2004: 9-14).

The results of Article 1 (and Article 2) come from a sample that is based on Dryer's (1989, 1992) g-sampling, modified for hypothesis-testing by Miestamo (2005: 31-39). This means that the results are generalizable to currently spoken languages and

thus represent a synchronic cross-linguistic trend. Although historical data also provide considerable evidence for a close relationship between dependent marking and rigid order, this correlation would have to be studied by investigating family-internal trends in order to give a more dynamic interpretation of the results.

ARTICLE 3: WORD ORDER IN ZERO-MARKING LANGUAGES. The correlation between zero marking and SVO word order has been proposed many times in the literature, but it has remained largely hypothetical, owing to the lack of proper typological testing (e.g., Vennemann 1973, 1974; Lehmann 1978; Blake 2001: 158; McWhorter 2001: 157-159; Wunderlich 2006: 95). Article 3 is the first research to substantiate this hypothesized correlation with a large database, testing the effect of the confounding factors in competition with structural pressure via multiple logistic regression.

The most important empirical result of Article 3 was thus to confirm with a large and balanced sample, and with multifactorial statistical tests, that a universal correlation exists between a morphological type of core argument marking, coded as zero vs. overt marking, and the dominant word order of the main clausal elements, S, O, and V, so that zero marking correlates with SVO order, and overt marking correlates with non-SVO order. This correlation was independent of the confounding areal and genealogical factors, regardless of how they were coded. The distribution of zero marking was further areally biased to Southeast Asia and to the Macro-Sudan Belt, an area that runs from West Africa to (roughly) the Western border of Uganda (Güldemann 2008), and it was genealogically biased to the Niger-Congo and Austro-Asiatic language families. The data were modeled by using a basically synchronic approach, namely, Dryer's (1992) g-sampling, together with an approach that allows more dynamic interpretations, namely, Bickel's (2008b, 2011) Family Bias Theory. The correlation between zero marking and SVO is thus, according to my definitions, both a synchronic and a dynamic universal (see Section 3.1.1).

In terms of language complexity, this article indicates that there is a positive correlation between morphological type and the syntactic organization of the clausal elements (see Section 2.5.3 for a detailed explanation of the simplicity of SVO order). That correlation happened to be in terms of simple morphology correlating with simple

syntactic property, but this correlation appears to be a rather special case in that the simplest choice (SVO order) was also the most distinctive, that is, the most economic and efficient in function. Based on earlier research by Fenk-Oczlon and Fenk (1999, 2008) and Maddieson (2006), one would expect that positive correlations were possible only within a coding domain, for instance, within phonology, and that only negative correlations would occur between different coding domains, such as between morphology and syntax. This expectation is cancelled out by the results of Article 3. Based on these results, I would hypothesize that any type of complexity correlation may be possible both within and across coding domains, but positive correlations are expected to cluster within a coding domain rather than across them; meanwhile across coding domains negative correlations are expected to be more frequent than positive correlations, owing to the fact that morphology and syntax tend to “counterbalance each other in sharing the cognitive-communicative load of language” (Givón 2009: 58). Demonstrating the truth of this hypothesis is a task for future research.

4.3.2. Question 2: Complexity in its geographical and sociocultural context

ARTICLE 2: COMPLEXITY IN CORE ARGUMENT MARKING AND POPULATION SIZE. In linguistics, it has often been assumed that language structure does not correlate with the geographical or sociocultural environment of the speech community (Sapir 1921; Kaye 1989: 48). Owing to the long-lasting influence of such assumptions, the relationship of language structure to its geographical and/or social environment has seldom been researched typologically. Article 2 is one of the first typological studies to test this assumption, and it provides empirical evidence that contradicts the widely held assumption.

The most important empirical result of Article 2 is that complexity in core argument marking correlates negatively with population size when complexity is measured as departures from the principle of one-meaning–one-form. More specifically, this result means that “small” languages tend to violate the principle of one-meaning–one-form more than “large” languages. Notable exceptions to this correlation were large languages violating economy, spoken mostly in the Old World. When these languages

were removed, the correlation was also independent of area, that is, a cross-linguistic trend in languages emerged. Several new tests were done in the Commentary on Article 2, based on some reanalyzed data, and these further corroborate the results.

This finding is not an isolated oddity, but connects to a wealth of recent research along similar lines (e.g., Trudgill 2004; Hay and Bauer 2007; Gil 2009; Nichols 2009; Lupyan and Dale 2010). It appears that, once linguists began to investigate the relationship between language structure and social environment in an empirical way, a number of very interesting results surfaced, which provide increasing evidence that social features may well affect language structure in various ways. Language structure is thus, at least to some extent, adaptive to its sociocultural (and geographical) environment. More research is needed on this very promising line of research.

OTHER ARTICLES. Although the geographical distribution of complexity was not a major research question in my work, the data provided evidence for a few areal patterns of complexity as well. For one thing, the distribution of zero marking was areally biased in a significant way, clustering in two areal hotbeds, the Macro-Sudan Belt and Southeast Asia. These areas have long been known to exhibit low levels of morphological complexity, but Article 3 shows this more clearly than before in the domain of core argument marking. One minor cluster of zero marking also occurred in Oaxaca, Mexico, where several verb-initial languages have no overt morphological marking of noun arguments. While the reason for zero marking to have clustered there remains unclear, in case of the two areal hotbeds it seems that language contact is the most likely cause for the development, spread, and maintenance of zero marking. An interesting further result was that the areal pattern of zero marking was so strong that it manifested regardless of how areas were coded in the tests. This means that one and the same variable can have an areally-biased distribution at multiple levels of granularity.

In Article 3, areal factors also had a much greater effect on the distribution of zero marking than word order. This result is somewhat unexpected from the viewpoint of more traditional typology, where the focus has been on language universals and constraints on possible language types, and less on areal and historical factors. Because the correlation between zero marking and SVO word order has been so often noted in

the literature, against this background it is a little surprising that areal factors in fact had a greater effect on the distribution of zero marking than did linguistic factors. However, as typologists have begun to pay more attention to areal and historical factors, it has become increasingly evident that the distribution of most typological variables is affected by areal factors (see especially Dryer 1989, 1992; Nichols 1992; Bickel 2007: 243-244). From this perspective, the strong effect of areal factors in Article 3 is unsurprising. On the other hand, the finding also further emphasizes the methodological need to separate areal factors from linguistic factors as well as the need to model typological distributions with multifactorial methods.

4.3.3. Relevance of different complexity types to language universals

Based on these empirical results, it appears that the different notions of complexity varied somewhat in their relevance to language universals (see Section 2.5 for a description of the article-wise complexity metrics). Below, I briefly discuss the relevance of Dahl's (2004) notions of complexity to language universals in my case studies.

First, system complexity was the most relevant to the correlation between linguistic structure and population size in Article 2. Second, structural complexity was most relevant to the correlations between the coding strategies in Articles 1 and 3. Third, while conceptual complexity was relevant to the correlation between linguistic structure and population size in Article 2, overall it did not play a major role in the case studies (see Sinnemäki, forthcoming, for a similar conclusion). As noted in Section 2.5, I also approached complexity trade-offs between the coding strategies from the viewpoint of conceptual complexity, but I did not include those results in Article 1, owing to the lack of any significant results in the pilot studies (Sinnemäki 2005, 2006).

Based on these observations, it seems that structural complexity was overall the most relevant to language universals in my case studies, while conceptual complexity was the least relevant. System complexity (especially in terms of form-meaning mapping) was the most relevant when correlating linguistic structure with population size. But why should structural complexity be more relevant to language universals than

the other two notions of complexity? One possible reason is that the structure of expressions is the most readily available for manipulation by speakers and also the most vulnerable to the effects of language use (such as frequency); by contrast, the conceptual level and the properties of the language's system may not be so easily affected by direct manipulation and/or the effects of language use (see also Good 2010).

All in all, the different notions of complexity defined by Dahl (2004) seemed to vary somewhat in their relevance to language universals. However, further research is needed to confirm these conclusions as well as to discover the reasons for them.

4.4. On explanations

In line with functional-typological research in general, I argue that the strongest explanations can be found among language-external factors, such as functional motivations, historical factors (e.g., language contacts), and the properties of the speech community (e.g., Givón 2001; Croft 2003; Hawkins 2004; Bickel 2007; Lupyan and Dale 2010). These issues are briefly discussed below.

In all of my case studies, I explained the results in terms of functional motivations, at least in part. In Article 1, the correlation between dependent marking and rigid order was explained as a matter of the interaction between the general principles of economy and distinctiveness. In Article 2, it was argued that processing preferences and language learning constraints favored adherence to the principle of one-meaning–one-form in large languages and allowed its violation in small languages. In Article 3, I argued that the correlation between zero marking and SVO word order could be explained by a combination of iconicity, economy, and processing preferences. Functional motivations were thus relevant for explaining many of the statistical patterns in my data.

Article 3 further showed that the distribution of zero marking was areally very biased in two areal hotbeds, the Macro-Sudan Belt and Southeast Asia. Language contact appears to be the most promising explanation for the emergence, spread, and maintenance of zero marking in these hotbed areas, but more research is needed to confirm this, especially in the case of the Macro-Sudan Belt. The distribution of zero

marking in these two hotbeds has not previously been researched in such detail as it is here, and, to my knowledge, the explanation in terms of language contact in these areas has also not been proposed earlier. Language contact was also raised in connection with zero marking in non-SVO languages, but I doubt whether that can explain all instances, such as the clustering of zero marking in Oaxaca, Mexico, discussed above.

The connection between population size and complexity is a more complicated correlation to explain, because these variables are not connected directly with one another, but probably via some intermediary factors, such as cognitive issues related to language learning (e.g., Gil 2009). In Article 2, I followed Trudgill (2004) and tentatively assumed that violations of economy and distinctiveness are affordable in small languages because of the generally low numbers of second-language learners, whereas large languages usually attract greater numbers of adult second-language learners, who tend to favor transparency and have difficulties with underspecification (see also Kusters 2003).

This idea has been advanced by Lupyan and Dale (2010), who suggest **the linguistic niche hypothesis** for explaining the close relationship between complexity and social environment. According to their proposal, languages adapt to the communicative needs and the learning constraints of the speaker population, which affect the level of morphological, lexical, and syntactic specification of languages. This approach helps to explain at a more general level why the complexity patterns in core argument marking were so strongly affected by the social environment. Population size has proven to be a useful parameter in typological research in modeling linguistic diversity and the rate of language change, for instance (see Nettle 1999; Wichmann, Stauffer, and Schulze 2008; Wichmann and Holman 2009). Although it is just one social factor, population size is a promising heuristic for estimating the effect of social environment at large.

In terms of causation, I argue that the statistical trends in core argument marking are best explained through multiple causation in that language structure adapts to the multiple forces of universal structural pressure and learning constraints in the speaker population. First, the interactions between the coding strategies in core argument marking appear to be caused by the functional motivations of economy and

distinctiveness on the one hand (Article 1), and by population size on the other hand (Article 2). Second, in Article 3, I showed with multiple logistic regression that the typological distribution of zero marking of core arguments was affected by structural factors, areal diffusion, and genealogical inheritance. It thus appears that none of the statistical trends could be fruitfully explained by recourse to a single explanation.

Although linguistic variables are often affected by numerous factors, I still do not consider multiple causation to be a default option. At its worst, too strong an adherence to multiple causation may mask ignorance of the mechanisms at work, and, therefore, multiple causation should be argued for only when sufficient evidence exists for doing so. Thus, although the nature of reality calls for multiple causation in most instances, we should not be lax in assessing the real contribution of each alleged explanation. One of the ways to assess the contribution of different factors is to use multifactorial modeling techniques for the purpose (e.g., Diessel 2008; Bickel 2008b).

Chapter 5 Conclusion

5.1. Main scholarly contribution

In my dissertation, I have studied language complexity from a typological perspective. Language is an immensely complex system, probably the most complex form of human behavior that exists (Locke 2008: 646). To begin modeling the complexity of language is not an easy task, but here some progress has been made toward that end. Meanwhile, I hope to have shown that the notion of complexity is relevant and useful to typological research. The main scholarly contribution of this work can be summarized as follows:

- A. I applied the notion of complexity to typology in a way that enabled cross-linguistic comparison and argued that five factors were key issues in this approach (in line with Dahl 2004 and Miestamo 2008): i) the separation of complexity from difficulty, ii) focusing on local instead of global complexity, iii) measuring complexity as description length of the structures in an entity (i.e., as effective complexity), iv) breaking up the notion of complexity into different types, and v) choosing comparanda whose complexity differences are rather large (see especially Chapter 2 and Article 1). My approach clarifies the notion of language complexity and showed its usefulness in the empirical studies; hopefully, it may serve as a catalyst for further research on language complexity.
- B. Evaluation of typological distributions was advanced by using three methodological developments: i) the insight of language universals as diachronic laws of type preference was implemented using Bickel's (2008b, 2011) Family Bias Theory, ii) the impact of multiple factors on typological distributions was estimated with logistic regression (Article 3, Commentary on Articles 1 and 2), and iii) the effect of confounding factors was modeled in multiple ways in the spirit of multi-methodology (e.g., Bickel 2008b; Arppe et al. 2010). This approach takes the evaluation of language universals forward.

- C. Five new empirical generalizations emerged from the data: i) a complexity trade-off occurred between dependent marking and rigid word order (Article 1), ii) head marking did not correlate in terms of complexity with the other coding strategies (Article 1), iii) complexity in core argument marking correlated negatively with population size (Article 2), iv) the zero marking of core arguments correlated universally with SVO word order (Article 3), and v) areal diffusion had a greater impact than word order on the distribution of zero marking (Article 3). Generalizations i), ii), and iii) were synchronic universals, while that of iv) was a dynamic universal. Generalizations i) and iv) verified two long-standing hypotheses, which have not previously been studied with the methods of quantitative typology, while Generalization iii) goes against long-held beliefs in the discipline. The different notions of complexity also varied in terms of their relevance to these generalizations, with structural complexity and system complexity being the most relevant (Dahl 2004).
- D. In terms of explanations, I elucidated the distribution of complexity by multiple causation: i) the interaction between the coding strategies could be affected by functional motivations (Article 1), but also by population size (Article 2) and ii) the distribution of zero marking was affected by word order, areal diffusion, and genealogical relatedness (Article 3). As in functional-typological research in general, I argued that the strongest explanations are language-external.

As for my two empirical research questions, the results provided evidence for systematic cross-linguistic variation in terms of complexity, and also some evidence for balancing effects of complexity in core argument marking. These effects were argued to be the result of the principles of economy and distinctiveness. To the extent that these principles are related to cognition (cf. Bornkessel-Schlesewsky and Schlewsky 2009), there may also be some evidence for cognitive mechanisms that could limit the complexity of particular structures (cf. Fenk-Oczlon and Fenk 1999; Miestamo 2008: 31-32; *pace* e.g., Shosted 2006). Yet such balancing effects were rather limited and cannot currently be held as an all-encompassing principle in language.

The results also provide evidence for assuming that linguistic patterns can be fruitfully linked with other anthropological variables, such as cognitive, geographical, and sociocultural patterns. If language structure was genetically fixed and just a matter of deep and abstract universal grammar, then it would make sense to assume that language complexity could not vary along geographical and sociocultural parameters. However, these assumptions do not seem plausible in the light of my results, which rather suggest that linguistic structure can be affected by language-external factors. All in all, my work provides evidence for language as a complex adaptive system, owing to the many interactions that linguistic patterns have among themselves and with the sociocultural properties of the speech community (cf. Ellis and Larsen-Freeman 2009; Section 4.1).

A final observation on my results is that the empirical generalizations presented here are fully falsifiable. They can be disproven by showing that the correlations are non-significant and/or dependent on confounding factors in the light of new and/or more reliable data. However, to the extent that the methods are warranted and the data are reliable, the results add to our understanding of human language and its complex structure.

5.2. Issues for further research

In this dissertation I studied linguistic patterns close to surface phenomena in the style of Greenbergian functional-typological research rather than on a deep and abstract level in the style of Chomskyan generative grammar. However, the typological variables of interest were studied at a reductionist level in that I did not intend to take into account the full variation of complexity in core argument marking, but rather coded the variables of interest binomially (as overt vs. non-overt marking) in most instances and limited the domain of inquiry to the most basic clausal configuration. These choices were motivated, on the one hand, by the fact that complexity metrics are the most useful when complexity differences between the comparanda are quite large (cf. Gell-Mann and Lloyd 2004), and on the other hand, by the need to make the workload manageable by limiting the degrees of freedom in the grammatical variation covered.

The practice of coding a limited amount of variation has been a general trend in studies on language universals. However, it has been recently challenged, for example, by Bickel (2007) and Dixon (2009: 257-263), who call for more analytical and detailed examination of typological variables. Bickel and Witzlack-Makarevich (2008) and Bickel, Hildebrand, and Schiering (2009) are two recent examples of how this view could be successfully implemented in the study of language universals (see also Section 3.2 for parallel text typology). These issues suggest that language complexity in core argument marking could and should be studied in the future with more analytical and detailed metrics. One way in which I am planning to take this challenge seriously is to study the variation in the semantic/pragmatic conditions that affect the differential case marking of arguments.

This variationist view of typology takes the field closer to other variationist subdisciplines of linguistics, especially those of sociolinguistics and dialectology. However, it is not yet clear what the consequences of this development are for the future of the study of language universals. My belief, based on my own results, the results of Bickel and Witzlack-Makarevich (2008), and the discussion I had with Peter Austin after his plenary talk (Austin 2009) at the conference on *Case In and Across Languages* in Helsinki on the 29th of August 2009, is that the more fully typological variation is coded, the more likely it is that the data will reflect areal and genealogical tendencies, that is, local grammaticalization paths, and the more often language universals will cluster around the more abstract or reductionist levels of coding typological variables. This is not to say that I expect universals to be found at the level of abstractness postulated in Universal Grammar, since generativists are often completely uninterested in the Greenbergian-style language universals accumulated during the past 50 years or so (e.g., Boeckx 2009). My opinion – and I may well be wrong – is that linguistic diversity is so overwhelmingly great that statistical universals are going to be exceedingly rare when coding typological variables at high levels of detail. This is not bad news for typology, since results showing local grammaticalization paths rather than universal tendencies advance our knowledge of cross-linguistic diversity to new levels and may enable their combination with other local patterns (see Bickel 2007).

There is also a growing awareness in linguistics that the direction for future research in the field is in multi-methodological approaches. This means that instead of emphasizing one particular method, different methods can and should be used for different kinds of data, especially to demonstrate converging evidence for the patterns found – in other words, to attempt replicating the results by using different data and methods. In the case of converging evidence, this makes the results less dependent on a particular method or technique. We should thus avoid relying too heavily on one method, such as intuition, corpus studies, or experiments, but rather acknowledge the advantages and limitations of each method and approach the research questions from multiple angles (e.g., Arppe and Järviö 2007; Haspelmath 2009; Arppe et al. 2010).

This is, of course, rarely possible within the limits of one publication produced by a single researcher. Instead multidisciplinary research programs that aim at joint publications around the same set of research questions are called for. For the study of language complexity, this suggests several issues for further research: i) trying out different coding of typological variables, ii) contrasting complexity metrics that pay attention to different types of complexity with the same variable of interest, and iii) studying the connection between complexity and difficulty by conducting psycholinguistic experiments. In effect, the crux of this line of research is that in order to gain deeper understanding of complex phenomena, the phenomena should be studied in multiple ways.

The issue of equal complexity of languages has come under much criticism in the last few years (e.g., McWhorter 2001; Kusters 2003; Maddieson 2006; Shosted 2006; Sampson 2009). I deliberately refrained from studying this claim and have in fact argued for the impossibility of studying it as a potential statistical universal. As mentioned in Chapter 2, I do not see how our understanding of human language would increase if we could scale languages according to global complexity. However, our understanding of human language would increase by studying the interconnections among different components of grammar as well as their cognitive bases. Shosted (2006), for one, found no evidence for such interconnections between phonology and morphology, and thus no evidence for cognitive mechanisms that could limit complexity. Although I claimed to have found some indirect evidence for such

mechanisms, I have had very little to say otherwise about cognitive issues related to complexity. Nevertheless, studying the cognitive basis of complexity correlations is an important area for further research (cf. Givón 2009: 11-14). As suggested in Chapter 4, the interaction between the general principles of economy and distinctiveness might offer a fruitful basis for such investigation, since these principles are relevant to language processing (e.g., Bornkessel-Schlesewsky and Schlewsky 2009).

The definition of complexity in this work differs somewhat from the notion of complexity used in the study of complex (adaptive) systems, a subfield of the science of complexity. In the study of complex adaptive systems, complexity is an emergent, non-linear property that cannot be reduced to the sum of its parts, owing to their numerous interactions. These properties increase the unpredictability of the system's behavior, necessitating holistic description (Rescher 1998: 26-27). To this end, complexity is not measured by attempting to define it objectively along several dimensions, as has been done here, but rather by "merely" validating whether or not a system fulfills certain properties known a priori (see Coupé, Marsico, and Pellegrino 2009: 141-142). Language has been approached from this viewpoint before, at least in the study of second language learning (Larsen-Freeman 1997; Ellis and Larsen-Freeman 2009), in phonology (Coupé, Marsico, and Pellegrino 2009; Pellegrino et al. 2009), and in semantics (Wildgen 2005). Yet, to my knowledge, typologists have not approached morphosyntax by treating it as a complex adaptive system, although this would no doubt open new ways of understanding the development and behavior of language structure from a typological perspective.

References

- Aboh, Enoch O. and Norval Smith 2009. Simplicity, simplification, complexity and complexification: Where have the interfaces gone? In Enoch O. Aboh and Norval Smith (eds.), *Complex processes in new languages*, 1-25. Amsterdam: John Benjamins.
- Anderson, Avery D. 2007. The major functions of the noun phrase. In Timothy Shopen (ed.), *Language typology and syntactic description, Volume I: Clause structure* (2nd edn), 132-223. Cambridge: Cambridge University Press.
- Ansaldo, Umberto and Sebastian N. Nordhoff 2009. Complexity and the age of languages. In Enoch O. Aboh and Norval Smith (eds.), *Complex processes in new languages*, 345-363. Amsterdam: John Benjamins.
- Ansaldo, Umberto and Stephen Matthews 2007. Deconstructing creole: The rationale. In Umberto Ansaldo, Stephen Matthews, and Lisa Lim (eds.), *Deconstructing creole*, 1-18. Amsterdam: John Benjamins.
- Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert and Arne Zeschel 2010. Cognitive Corpus Linguistics: Five points of debate on current theory and methodology. *Corpora* 5(1): 1-27.
- Arppe, Antti and Juhani Järvikivi 2007. Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory* 3(1): 99-109.
- Austin, Peter 2009. Issues in the synchrony and diachrony of case marking in the Mantharta and Kanyara languages, Australia. Plenary presentation at the conference *Case in and across languages*, Helsinki, August 29, 2009.
- Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall, and Ingo Plag 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* (Special issue on Creoles and typology, ed. by Parth Bhatt and Tonjes Veenstra) 26(1): 5-42. <http://www2.uni-siegen.de/~engspra/>. (Accessed November 21, 2010).
- Bane, Max 2008. Quantifying and measuring morphological complexity. In Charles B. Chang and Hannah J. Haynie (eds.), *Proceedings of the 26th West Coast*

- conference on formal linguistics*, 69-76. Somerville, MA: Cascadilla Proceedings Project.
- Battistella, Edwin L. 1990. *Markedness: The evaluative superstructure of language*. Albany: State University of New York Press.
- Bell, Alan 1978. Language samples. In Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of human languages, Volume 1: Method & theory*, 123-156. Stanford: Stanford University Press.
- Bichakjian, Bernard H. 1999. Language evolution and the complexity criterion. *Psychology* 10(33).
- Bickel, Balthasar 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1): 239-251.
- Bickel, Balthasar 2008a. A refined sampling procedure for genealogical control. *Language Typology and Universals (STUF)* 61(3): 221-233.
- Bickel, Balthasar 2008b. A general method for the statistical evaluation of typological distributions. Manuscript, University of Leipzig. <http://www.uni-leipzig.de/~bickel/>. (Accessed April 17, 2008.)
- Bickel, Balthasar 2010. Absolute and statistical universals. In Patrick Hogan (ed.), *The Cambridge encyclopedia of language sciences*, 77-79. Cambridge: Cambridge University Press.
- Bickel, Balthasar 2011. Distributional biases in language families. Manuscript, University of Zurich, revised version. <http://www.uzh.ch/spw/aboutus/bickel/>. (Accessed July 21, 2011.)
- Bickel, Balthasar and Alena Witzlack-Makarevich 2008. Referential scales and case alignment: Reviewing the typological evidence. In Marc Richards and Andrej L. Malchukov (eds.), *Scales*, 1-37. Leipzig: University of Leipzig.
- Bickel, Balthasar and Johanna Nichols 2006. Oceania, the Pacific Rim and the theory of linguistic areas. *Proceedings of the thirty-second annual meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society.
- Bickel, Balthasar and Johanna Nichols 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic description, Volume III:*

- Grammatical categories and the lexicon* (2nd edn), 169-240. Cambridge: Cambridge University Press.
- Bickel, Balthasar, Kristine A. Hildebrandt, and René Schiering 2009. The distribution of phonological word domains: A probabilistic typology. In Janet Grijzenhout and Barış Kabak (eds.), *Phonological domains: Universals and deviations*, 47-75. Berlin: Mouton de Gruyter.
- Bickerton, Derek 1995. *Language and human behaviour*. London: UCL Press.
- Bisang, Walter 2004a. Dialectology and typology - An integrative perspective. In Bernd Kortmann (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*, 11-45. Berlin: Mouton de Gruyter.
- Bisang, Walter 2004b. Reproducibility in linguistics? Manuscript, University of Mainz. <http://www.linguistik.uni-mainz.de/bisang/publications/>. (Accessed November 11, 2009.)
- Bisang, Walter 2009. On the evolution of complexity: Sometimes less is more in East and mainland Southeast Asia. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*, 34-49. Oxford: Oxford University Press.
- Blake, Barry J. 2001. *Case* (2nd edn). Cambridge: Cambridge University Press.
- Boeckx, Cedric 2009. Contribution to Round table on linguistic universals. In Massimo Piatelli-Palmarini, Juan Uriagereka, and Pello Salaburu (eds.), *Of minds and language: The Basque country encounter with Noam Chomsky*, 195-220. Oxford: Oxford University Press.
- Bornkessel-Schlesewsky, Ina and Matthias Schlewsky 2009. Minimality as vacuous distinctness: Evidence from cross-linguistic sentence comprehension. *Lingua* 119(10): 1541-1559.
- Bowern, Claire 2009. Defining complexity: Historical reconstruction and Nyulnyulan subordination. *Rice Working Papers in Linguistics* 1: 1-17.
- Brazzale, Alessandra R. and Anthony C. Davison 2008. Accurate parametric inference for small samples. *Statistical Science* 23(4): 465-484.
- Breiman, Leo 2001. Random forests. *Machine Learning* 45(1): 5-32.
- Brown, Lea 2001. A grammar of Nias Selatan. Ph.D. diss., University of Sydney.

- Bybee, Joan L. and Sandra A. Thompson 1997. Three frequency effects in syntax. *Proceedings of the twenty-third annual meeting of the Berkeley Linguistics Society*, 65-85. Berkeley: Berkeley Linguistics Society.
- Chaitin, Gregory J. 1987. *Algorithmic information theory*. Cambridge: Cambridge University Press.
- Chomsky, Noam 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam and Morris Halle 1968. *Sound pattern of English*. New York: Harper and Row.
- Comrie, Bernard 1992. Before complexity. In John A. Hawkins and Murray Gell-Mann (eds.), *The evolution of human languages: Proceedings of the workshop on the Evolution of human languages, held August, 1989 in Santa Fe, New Mexico*, 193-211. Redwood City: Addison-Wesley.
- Comrie, Bernard 2005. Alignment of case marking. In Martin Haspelmath et al. (eds.), *The world atlas of language structures*, 398-405. Oxford: Oxford University Press.
- Corbett, Greville G. 2005. Suppletion in personal pronouns: Theory versus practice, and the place of reproducibility in typology. *Linguistic Typology* 9(1): 1-23.
- Coupé, Christophe, Egidio Marsico, and François Pellegrino 2009. Structural complexity of phonological systems. In François Pellegrino, Egidio Marsico, Ioana Chitoran, and Christophe Coupé (eds.), *Approaches to phonological complexity*, 141-169. Berlin: Mouton de Gruyter.
- Croft, William 2000. *Explaining language change: An evolutionary approach*. Harlow: Pearson Education.
- Croft, William 2003. *Typology and universals* (2nd edn). Cambridge: Cambridge University Press.
- Croft, William and D. Alan Cruse 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Crystal, David 1997. *The Cambridge encyclopedia of language* (2nd edn). Cambridge: Cambridge University Press.

- Cysouw, Michael 2007. New approaches to cluster analysis of typological indices. In Peter Grzybek and Reinhard Köhler (eds.), *Exact methods in the study of language and text: Dedicated to Gabriel Altmann on the occasion of his 75th birthday*, 61-76. Berlin: Mouton de Gruyter.
- Cysouw, Michael 2010. Dealing with diversity: Towards an explanation of NP word order frequencies. *Linguistic Typology* 14(2): 253-287.
- Cysouw, Michael and Bernhard Wälchli (eds.) 2008. Parallel texts. Using translational equivalents in linguistic typology. Theme issue in *Language Typology and Universals (STUF)* 60(2): 95-181.
- Dahl, Östen 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- Dahl, Östen 2008. Grammatical resources and linguistic complexity: Sirionó as a language without NP coordination. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 153-163. Amsterdam: John Benjamins.
- Dahl, Östen 2009. Testing the assumption of complexity invariance: The case of Elfdalian and Swedish. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*, 50-63. Oxford: Oxford University Press.
- DeGraff, Michel 2001. On the origin of creoles: A Cartesian critique of Neo-Darwinian linguistics. *Linguistic Typology* 5(2/3): 213-310.
- Deutscher, Guy 2006. *The unfolding of language: An evolutionary tour of mankind's greatest invention*. London: William Heinemann.
- Deutscher, Guy 2009. 'Overall complexity' – a wild goose chase? In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*, 243-251. Oxford: Oxford University Press.
- Diessel, Holger 2008. Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics* 19(3): 465-490.
- Diller, Anthony 1993. Diglossic grammaticality in Thai. In William A. Foley (ed.), *The role of theory in language description*, 393-420. Berlin: Mouton de Gruyter.

- Dixon, R.M.W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Dixon, R.M.W. 2009. *Basic linguistic theory Volume 1: Methodology*. Oxford: Oxford University Press.
- Dixon, R.M.W. 2010. *Basic linguistic theory Volume 2: Grammatical topics*. Oxford: Oxford University Press.
- Doebeli, Michael and Iaroslav Ispolatov 2010. Complexity and diversity. *Science* 328(5977): 494-497.
- Dol, Philomena 1999. A grammar of Maybrat: A language of the Bird's Head, Irian Jaya, Indonesia. Ph.D. diss., Leiden University.
- Dressler, Wolfgang U. 2003. Naturalness and morphological change. In Brian D. Joseph and Richard D. Janda (eds.), *The handbook of historical linguistics*, 461-471. Oxford: Blackwell.
- Dressler, Wolfgang U., Willi Mayerthaler, Oswald Panagl, and Wolfgang U. Wurzel 1987. *Leitmotifs in Natural Morphology*. Amsterdam: John Benjamins.
- Dryer, Matthew 1989 Large linguistic areas and language sampling. *Studies in Language* 13(2): 257-292.
- Dryer, Matthew 1992. The Greenbergian word order correlations. *Language* 68(1): 81-138.
- Dryer, Matthew 1997. On the six-way word order typology. *Studies in Language* 21(1): 69-103.
- Dryer, Matthew 1998. Why statistical universals are better than absolute universals. *Papers from the 33rd annual meeting of the Chicago Linguistic Society (CLS)* 33(2): 123-145.
- Dryer, Matthew 2000. Counting genera vs. counting languages: A reply to Maslova. *Linguistic Typology* 4(3): 334-350.
- Dryer, Matthew 2002. Case distinctions, rich verb agreement, and word order type. *Theoretical Linguistics* 28(1): 151-157.
- Dryer, Matthew 2005a. Genealogical language list. In Martin Haspelmath et al. (eds.), *The world atlas of language structures*, 584-644. Oxford: Oxford University Press.

- Dryer, Matthew 2005b. Order of subject, object, and verb. In Martin Haspelmath et al. (eds.), *The world atlas of language structures*, 330-333. Oxford: Oxford University Press.
- Dryer, Matthew 2006. Descriptive theories, explanatory theories, and basic linguistic theory. In Felix K. Ameka, Alan Charles Dench, and Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 207-234. Berlin: Mouton de Gruyter.
- Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley, and Stephen C. Levinson 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309: 2072-2075.
- Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink, and Angela Terrill 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84(4): 710-759.
- Edmonds, Bruce 1999. Syntactic measures of complexity. Ph.D. diss., University of Manchester.
- Ellis, Nick C. and Diane Larsen-Freeman (eds.). Language as a complex adaptive system. *Language Learning* 59, Supplement 1.
- van Enk, Gerrit J. and Lourens de Vries 1997. *The Korowai of Irian Jaya: Their language in its cultural context*. Oxford: Oxford University Press.
- Evans, Nicholas and Stephen C. Levinson 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5): 429-492.
- Everett, Daniel L. 1986. Pirahã. In Desmond C. Derbyshire and Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages*, vol. 1, 200-325. Berlin: Mouton de Gruyter.
- Everett, Daniel L. 1987. Pirahã clitic doubling. *Natural Language and Linguistic Theory* 5(2): 245-276.
- Fast, Lesley 1990. *Tungak grammar essentials*. In John R. Roberts (ed.), *Two grammatical studies*, 1-72. Ukarumpa: Summer Institute of Linguistics.
- Fenk-Oczlon, Gertraud and August Fenk 1999. Cognition, quantitative linguistics, and systemic typology. *Linguistic Typology* 3(2): 151-177.

- Fenk-Oczlon, Gertraud and August Fenk 2008. Complexity trade-offs between the subsystems of language. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 43-65. Amsterdam: John Benjamins.
- Fodor, Janet Dean 1998. Learning to parse. *Journal of Psycholinguistic Research* 27(2): 285-319.
- Fortescue, Michael 1984. *West Greenlandic*. London: Croom Helm.
- Frajzyngier, Zygmunt and Erin Shay 2003. *Explaining language structure through systems interaction*. Amsterdam: John Benjamins.
- Frank, Robert 2004. Restricting grammatical complexity. *Cognitive Science* 28(5): 669-697.
- Gell-Mann, Murray 1994. *The quark and the jaguar: Adventures in the simple and the complex*. New York: W.H. Freeman.
- Gell-Mann, Murray 1995. What is complexity? *Complexity* 1(1): 16-19.
- Gell-Mann, Murray and Seth Lloyd 2004. Effective complexity. In Murray Gell-Mann and Constantino Tsallis (eds.), *Nonextensive entropy – interdisciplinary applications*, 387-398. Oxford: Oxford University Press.
- Gil, David 2005. Word order without syntactic categories. How Riau Indonesian does it. In Andrew Carnie, Heidi Harley, and Sheila Ann Dooley (eds.), *Verb first: On the syntax of verb-initial languages*, 243-263. Amsterdam: John Benjamins.
- Gil, David 2008. How complex are isolating languages. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 109-131. Amsterdam: John Benjamins.
- Gil, David 2009. Population size and the encoding of thematic roles. Paper presented at *The eight biennial meeting of the Association for Linguistic Typology (ALT 8)*, Berkeley, July 24, 2009.
- Givón, Talmy 1979. *On understanding grammar*. New York: Academic Press.
- Givón, Talmy 1981. Typology and functional domains. *Studies in Language* 5(2): 163-193.
- Givón, Talmy 2001. *Syntax: An introduction* (vol. 1). Amsterdam: John Benjamins.

- Givón, Talmy 2009. *The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam: John Benjamins.
- Good, Jeff 2010. Typologizing grammatical complexities *or* Why creoles may be paradigmatically simple but syntagmatically average. To appear in *Journal of Pidgin and Creole Languages*. <http://www.acsu.buffalo.edu/~jcgood>. (Accessed November 27, 2010.)
- Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language* (2nd edn), 73-113. Cambridge, MA: MIT Press.
- Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. In Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of human language I: Method and theory*, 61-92. Stanford: Stanford University Press.
- Greenberg, Joseph H. 1995. The diachronic typological approach to language. In Masayoshi Shibatani and Theodora Bynon (eds.), *Approaches to language typology*, 145-166. Oxford: Clarendon Press.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum.
- Güldemann, Tom 2008. The Macro-Sudan belt: Towards identifying a linguistic area in northern Sub-Saharan Africa. In Bernd Heine and Derek Nurse (eds.), *A linguistic geography of Africa*, 151-185. Cambridge: Cambridge University Press
- Guirardello, Raquel 1999. A reference grammar of Trumai. Ph.D. diss., Rice University.
- Haiman, John 1983. Iconic and economic motivation. *Language* 59(4): 781-819.
- Harrell, Frank E. Jr. 2010. *Design: Design package* (R package version 2.3-0). <http://biostat.mc.vanderbilt.edu/rms>. (Accessed March 6, 2010.)
- Harrell, Frank E. Jr. et al. 2010. *Hmisc: Harrell miscellaneous* (R package version 3.7-0). <http://CRAN.R-project.org/package=Hmisc>. (Accessed March 6, 2010.)
- Harris, Alice C. 2008. On the explanation of typologically unusual structures. In Jeff Good (ed.), *Linguistic universals and language change*, 54-76. Oxford: Oxford University Press.

- Harris, Alice, Larry M. Hyman, and James V. Staros 2006. What is reproducibility? *Linguistic Typology* 10(1): 69-73.
- Haspelmath, Martin 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1): 25-70.
- Haspelmath, Martin 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1): 1-33.
- Haspelmath, Martin 2009. Welche Fragen können wir mit herkömmlichen Daten beantworten? *Zeitschrift für Sprachwissenschaft* 28(1): 157-162.
- Haspelmath, Martin 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3): 663-687.
- Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Haspelmath, Martin and Sven Siegmund 2006. Simulating the replication of some of Greenberg's word order generalizations. *Linguistic Typology* 10(1): 74-82.
- Hauser, Marc D., Noam Chomsky and W. Tecumseh Fitch 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598): 1569-1579.
- Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hawkins, John A. 2009. An efficiency theory of complexity and related phenomena. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*, 252-268. Oxford: Oxford University Press.
- Hay, Jennifer and Laurie Bauer 2007. Phoneme inventory size and population size. *Language* 83(2): 388-400.
- Hengeveld, Kees, Jan Rijkhoff, and Anna Siewierska 2004. Parts-of-speech systems and word order. *Journal of Linguistics* 40(3): 527-570.
- Herrmann, Stephanie 2004. Warao. In Philipp Strazny (ed.), *Encyclopedia of linguistics*, vol. 2. London: Routledge.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.

- Hornby, Albert Sydney and Sally Wehmeier (compilers) 2007. *Oxford advanced learner's dictionary* (7th edn). Oxford: Oxford University Press. <http://www.oxfordadvancedlearnersdictionary.com/>. (Accessed November 25, 2010.)
- Huberman, Bernardo A. and Tad Hogg 1986. Complexity and adaptation. *Physica D* 22(1-3): 376-384.
- van der Hulst, Harry (ed.) 2010. *Recursion and human language*. Berlin: Mouton de Gruyter.
- Hübler, Alfred W. 2007. Understanding complex systems. *Complexity* 12(5): 9-11.
- Janssen, Dirk, Balthasar Bickel, and Fernando Zúñiga 2006. Randomization tests in language typology. *Linguistic Typology* 10(3): 419-440.
- Juola, Patrick 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics* 5(3): 206-213.
- Juola, Patrick 2008. Assessing linguistic complexity. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 89-108. Amsterdam: John Benjamins.
- Karlsson, Fred, Matti Miestamo, and Kaius Sinnemäki 2008. Introduction: The problem of language complexity. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, vii-xiv. Amsterdam: John Benjamins.
- Kaye, Jonathan 1989. *Phonology: A cognitive view*. Hillsdale, NJ: Lawrence Erlbaum.
- Keller, Rudi 1994. *On language change: The invisible hand in language*. London: Routledge.
- Kertész, András and Csilla Rákosi 2008. Conservatism vs. innovation in the (un)grammaticality debate. In András Kertész and Csilla Rákosi (eds.), *New approaches to linguistic evidence: Pilot studies*, 61-84. Frankfurt am Main: Peter Lang.
- Kiparsky, Paul 1968. Linguistic universals and linguistic change. In Emmon Bach and Robert T. Harms (eds.), *Universals in linguistic theory*, 170-202. New York: Holt, Rinehart and Winston.

- Kiparsky, Paul 1997. The rise of positional licensing. In Ans van Kemenade and Nigel Vincent (eds.), *Parameters of morphosyntactic change*, 460-494. Cambridge: Cambridge University Press.
- Koptjevskaja-Tamm, Maria 2010. Linguistic typology and language contact. In Jae Jung Song (ed.) *The Oxford handbook of linguistic typology*, 568-590. Oxford: Oxford University Press.
- Kortmann, Berndt 1996. *Adverbial subordination: A typology and history of adverbial subordinators based on European languages*. Berlin: Mouton de Gruyter.
- Kusters, Wouter 2003. Linguistic complexity: The influence of social change on verbal inflection. Ph.D. diss., University of Leiden.
- Kusters, Wouter 2008. Complexity in linguistic theory, language learning and language change. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 3-22. Amsterdam: John Benjamins.
- LaPolla, Randy J. 2005. Typology and complexity. In James W. Minett and William S-Y. Wang (eds.), *Language acquisition, change and emergence: Essays in evolutionary linguistics*, 465-493. Hong Kong: City University of Hong Kong Press.
- Larsen-Freeman, Diane 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics* 18(2): 141-165.
- Lehmann, Winfred P. 1978. English: A characteristic SVO language. In Winfred P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 169-222. Austin: University of Texas Press.
- Levinson, Stephen C. 2006. The language of space in Yélî Dnye. In Stephen C. Levinson and David P. Wilkins (eds.), *Grammars of space: Explorations in cognitive diversity*, 157-205. Cambridge: Cambridge University Press.
- Levinson, Stephen C. 2007. Optimizing person reference - Perspectives from usage on Rossel Island. In Nick J. Enfield and Tanya Stivers (eds.), *Person reference in interaction*, 29-72. Cambridge: Cambridge University Press.
- Li, Ming and Paul M.B. Vitányi 2008. *An introduction to Kolmogorov complexity and its applications* (3rd edn). New York: Springer.

- Lloyd, Seth 2001. Measures of complexity: A nonexhaustive list. *IEEE Control Systems Magazine* 21(4): 7-8.
- Locke, John L. 2008. Cost and complexity: Selection for speech and language. *Journal of Theoretical Biology* 251(4): 640-652.
- Longa, Víctor M. 2001. Sciences of complexity and language origins: An alternative to natural selection. *Journal of Literary Semantics* 30(1): 1-17.
- Lupyan, Gary and Rick Dale 2010. Language structure is partly determined by social structure. *PLoS One* 5(1).
- Mackenzie, Lachlan J. 2009. Content interrogatives in a sample of 50 languages. *Lingua* 119(8): 1131-1163.
- Maddieson, Ian 2006. Correlating phonological complexity: Data and validation. *Linguistic Typology* 10(1): 106-123.
- Maslova, Elena 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4(3): 307-333.
- Maslova, Elena and Tatiana Nikitina 2007. Stochastic universals and dynamics of cross-linguistic distributions: the case of alignment types. Manuscript, Stanford University. <http://anothersumma.net/>. (Accessed September 7, 2010.)
- Mayerthaler, Willi 1987. System-independent morphological naturalness. In Wolfgang U. Dressler, Willi Mayerthaler, Oswald Panagl, and Wolfgang U. Wurzel, *Leitmotifs in Natural Morphology*, 25-58. Amsterdam: John Benjamins.
- McAllister, James W. 2003. Effective complexity as a measure of information content. *Philosophy of Science* 70(2): 302-307.
- McWhorter, John 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2/3): 125-156.
- McWhorter, John 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.
- Meyer, David, Achim Zeileis, and Kurt Hornik 2006. The strucplot framework: Visualizing multi-way contingency tables with vcd. *Journal of Statistical Software* 17(3): 1-48.

- Meyer, David, Achim Zeileis, and Kurt Hornik 2010. *Vcd: Visualizing categorical data* (R package version 1.2-8). <http://cran.r-project.org/web/packages/vcd/vcd.pdf>. (Accessed March 6, 2010.)
- Miestamo, Matti 2005. *Standard negation: The negation of declarative verbal main clauses in a typological perspective*. Berlin: Mouton de Gruyter.
- Miestamo, Matti 2006. On the feasibility of complexity metrics. In Krista Kerge and Maria-Maren Sepper (eds.), *Finest linguistics. Proceedings of the annual Finnish and Estonian conference of linguistics, Tallinn, May 6-7, 2004*, 11-26. Tallinn: Tallinn University Press.
- Miestamo, Matti 2007. Symmetric and asymmetric encoding of functional domains, with remarks on typological markedness. In Matti Miestamo and Bernhard Wälchli (eds.), *New challenges in typology: Broadening the horizons and redefining the foundations*, 293-314. Berlin: Mouton de Gruyter.
- Miestamo, Matti 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 23-41. Amsterdam: John Benjamins.
- Miestamo, Matti 2009. Implicational hierarchies and grammatical complexity. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*, 80-97. Oxford: Oxford University Press.
- Miestamo, Matti, Kaius Sinnemäki, and Fred Karlsson (eds.) 2008. *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.
- Moravcsik, Edith A. and Jessica R. Wirth 1986. Markedness: An overview. In Fred R. Eckman, Edith A. Moravcsik, and Jessica R. Wirth (eds.), *Markedness*, 1-11. New York: Plenum.
- Moscoso del Prado Martín, Fermín 2010. The effective complexity of language: English requires at least an infinite grammar. Manuscript submitted to *Cognition*. <http://www.moscodelprado.net>. (Accessed November 4, 2010.)
- Moscoso del Prado Martín, Fermín 2011. The universal shape of human languages: Spectral analysis beyond speech. Manuscript, CNRS Lyon & Marseille. <http://www.moscodelprado.net>. (Accessed July 21, 2011.)
- Nettle, Daniel 1999. *Linguistic diversity*. Oxford: Oxford University Press.

- Newmeyer, Frederick J. 1986. *The politics of linguistics*. Chicago: The University of Chicago Press.
- Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, MA: MIT Press.
- Newmeyer, Frederick J. 2007. More complicated and hence, rarer: A look at grammatical complexity and crosslinguistic rarity. In Simin Karimi, Vida Samiian, and Wendy K. Wilkins (eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation*, 221-242. Berlin: Mouton de Gruyter.
- Nichols, Johanna 1986. Head-marking and dependent-marking grammar. *Language* 62(1): 56-119.
- Nichols, Johanna 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.
- Nichols, Johanna 2007. What, if anything, is typology? *Linguistic Typology* 11(1): 231-238.
- Nichols, Johanna 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*, 110-125. Oxford: Oxford University Press.
- Nichols, Johanna and Balthasar Bickel 2009. *The Autotyp genealogy and geography database: 2009 release*. <http://www.uni-leipzig.de/~autotyp/available.html>. (Accessed March 10, 2010.)
- Nieminen, Lea 2007. A complex case: A morphosyntactic approach to complexity in early child language. Ph.D. diss., University of Jyväskylä.
- Osborn, Henry A. Jr. 1966. Warao II: Nouns, relationals, and demonstratives. *International Journal of American Linguistics* 32(3): 253-261.
- Øvrelid, Lilja 2004. Disambiguation of syntactic functions in Norwegian. In Fred Karlsson (ed.), *Proceedings of the 20th Scandinavian Conference of Linguistics, Helsinki, January 7-9, 2004*. Helsinki: University of Helsinki, Department of General Linguistics. <http://www.ling.helsinki.fi/kielitiede/20scl/>. (Accessed December 1, 2007.)
- Page, Scott E. 2011. *Diversity and complexity*. Princeton: Princeton University Press.

- Palmer, Frank R. 1994. *Grammatical roles and relations*. Cambridge: Cambridge University Press.
- Parkvall, Mikael 2008. The simplicity of creoles in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 265-285. Amsterdam: John Benjamins.
- Payne, Doris L. (ed.) 1992. *Pragmatics of word order flexibility*. Amsterdam: John Benjamins.
- Peduzzi, Peter, John Concato, Elizabeth Kemper, Theodore R. Holford, and Alvan R. Feinstein 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49(12): 1373-1379.
- Pellegrino, François, Egidio Marsico, Ioana Chitoran, and Christophe Coupé 2009. Introduction. In François Pellegrino, Egidio Marsico, Ioana Chitoran, and Christophe Coupé (eds.), *Approaches to phonological complexity*, 1-17. Berlin: Mouton de Gruyter.
- Pellegrino, François, Egidio Marsico, Ioana Chitoran, and Christophe Coupé (eds.) 2009. *Approaches to phonological complexity*. Berlin: Mouton de Gruyter.
- Pericliev, Vladimir 2004. There is no correlation between the size of a community speaking a language and the size of the phonological inventory of that language. *Linguistic Typology* 8(3): 376-383.
- Perkins, Revere D. 1989. Statistical techniques for determining language sample size. *Studies in Language* 13(2): 293-315.
- Perkins, Revere D. 1992. *Deixis, grammar, and culture*. Amsterdam: John Benjamins.
- Pinker, Steven and Paul Bloom 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13(4): 707-784.
- Popper, Karl R. 1959. *The logic of scientific discovery*. New York: Basic Books.
- Prideaux, Gary D. 1970. On the selection problem. *Research on Language & Social Interaction* 2(2): 238-266.
- Primus, Beatrice 1999. *Cases and thematic roles: Ergative, accusative and active*. Tübingen: Niemeyer.

- R Development Core Team 2010. *R: A language and environment for statistical computing* (R version 2.10.1). Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org>. (Accessed March 6, 2010.)
- Rescher, Nicholas 1998. *Complexity: A philosophical overview*. New Brunswick: Transaction.
- Rijkhoff, Jan and Dik Bakker 1998. Language sampling. *Linguistic Typology* 2(3): 263-314.
- Rijkhoff, Jan, Dik Bakker, Kees Hengeveld, and Peter Kahrel 1993. A method for language sampling. *Studies in Language* 17(1): 169-203.
- Ritter, Nancy A. 2005. On the status of linguistics as cognitive science. *The Linguistic Review* 22(2-4): 117-133.
- Sampson, Geoffrey 2007. Grammar without grammaticality. *Corpus Linguistics and Linguistic Theory* 3(1): 1-32.
- Sampson, Geoffrey 2009. A linguistic axiom challenged. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*, 1-18. Oxford: Oxford University Press.
- Sampson, Geoffrey, David Gil, and Peter Trudgill (eds.) 2009. *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Sapir, Edward 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Company.
- Schaub, William 1985. *Babungo*. London: Croom Helm.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27: 379-423, 623-656.
- Shosted, Ryan K. 2006. Correlating complexity: A typological approach. *Linguistic Typology* 10(1): 1-40.
- Siewierska, Anna 1998. Variation in major constituent order: A global and a European perspective. In Anna Siewierska (ed.), *Constituent order in the languages of Europe*, 475-551. Berlin: Mouton de Gruyter.
- Siewierska, Anna (ed.) 1998. *Constituent order in the languages of Europe*. Berlin: Mouton de Gruyter.

- Simon, Herbert A. 1996. *The sciences of the artificial* (3rd edn). Cambridge, MA: MIT Press.
- Sinnemäki, Kaius 2005. Complex relationships between case marking, agreement and word order. Presentation at the conference *Approaches to Complexity in Language*, Helsinki, August 25, 2005.
- Sinnemäki, Kaius 2006. Compensations in the marking of core arguments. Presentation at the *Leipzig Students' Conference in Linguistics*, Leipzig, March 25, 2006.
- Sinnemäki, Kaius (forthcoming). Differential and consistent case marking of object: A typological study. A special issue of *Linguistics*, guest-edited by Giorgio Iemmolo and Gerson Klumpp.
- Sinnemäki, Kaius (in preparation). Cross-linguistic complexity correlations: Scope, nature and cause. Manuscript in preparation, University of Helsinki.
- Somers, Robert H. 1962. A new asymmetric measure of association for ordinal variables. *American Sociological Review* 27(6): 799-811.
- Song, Jae Jung 2006. What or where can we do better? Some personal reflections on (the tenth anniversary of) Linguistic Typology. *Linguistic Typology* 11(1): 5-22.
- Surendran, Dinoj and Partha Niyogi 2006. Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In Ole Nedergaard Thomsen (ed.), *Competing models of linguistic change: Evolution and beyond*, 43-58. Amsterdam: John Benjamins.
- Trudgill, Peter 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8(3): 305-320.
- van Valin, Robert Jr. 1977. Aspects of Lakhota syntax: A study of Lakhota (Teton Dakota) syntax and its implications for Universal Grammar. Ph.D. diss., University of California, Berkeley.
- Vennemann, Theo 1973. *Explanation in syntax*. In John Kimball (ed.), *Syntax and semantics* (vol. 2), 1-50. New York: Academic Press.
- Vennemann, Theo 1974. Topics, subjects, and word order: From SXV to SVX via TVX. In John Anderson and Charles Jones (eds.), *Historical linguistics: Proceedings of the first international congress of historical linguistics, Edinburgh, September 1973* (vol. II), 339-376. Amsterdam: North-Holland.

- Vilkki, Liisa (forthcoming). Semantic typology of grammatically expressed epistemic modality and inferentiality. Ph.D. diss., University of Helsinki.
- Vulanovic, Relja 2007. On measuring language complexity as relative to the conveyed linguistic information. *SKY Journal of Linguistics* 20: 399-437.
- Wälchli, Bernhard 2007. Constructing semantic maps from parallel text data. Manuscript, University of Konstanz. http://ling.uni-konstanz.de/pages/home/a20_11/waelchli/. (Accessed November 22, 2010.)
- Wichmann, Søren, Dietrich Stauffer, Christian Schulze, and Eric W. Holman 2008. Do language change rates depend on population size? *Advances in Complex Systems* 11(3): 357-369.
- Wichmann, Søren and Eric W. Holman 2009. Population size and rates of language change. *Human Biology* 81(2-3): 259-274.
- Widmann, Thomas and Peter Bakker 2006. Does sampling matter? *Linguistic Typology* 10(1): 83-95.
- Wildgen, Wolfgang 2005. Catastrophe theoretical models in semantics. In Reinhard Köhler, Gabriel Altmann, and Raïmond Genrikhovich Piotrovskii (eds.), *Internationales Handbuch der quantitativen Linguistik / International handbook of quantitative linguistics*, 410-423. Berlin: Mouton de Gruyter.
- Wojdak, Rachel 2005. The linearization of affixes: Evidence from Nuu-chah-nulth. Ph.D. diss, University of British Columbia.
- Wunderlich, Dieter 2006. Towards a structural typology of verb classes. In Dieter Wunderlich (ed.), *Advances in the theory of the lexicon*, 57-166. Berlin: Mouton de Gruyter.
- Wurzel, Wolfgang Ullrich 2001. Creoles, complexity, and linguistic change. *Linguistic Typology* 5(2/3): 377-387.
- Zamar, David, Brian McNeney, and Jinko Graham 2007. elrm: Software implementing exact-like inference for logistic regression models. *Journal of Statistical Software* 21(3): 1-18.
- Zipf, George K. 1948. *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley.

PART II Articles

List of Articles

Article 1: Kaius Sinnemäki 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language complexity: Typology, contact, change*. Studies in Language Companion Series 94, 67-88. Amsterdam/Philadelphia: John Benjamins.

Article 2: Kaius Sinnemäki 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language complexity as an evolving variable*. Studies in the Evolution of Language 13, 126-140. Oxford: Oxford University Press.

Article 3: Kaius Sinnemäki 2010. Word order in zero-marking languages. *Studies in Language* 34(4): 869-912.

The articles are included here as part of the dissertation with permission from John Benjamins Publishing Company and Oxford University Press.

ARTICLE 1:
Complexity trade-offs in core argument marking

Commentary on Article 1

As mentioned in Section 1.3 of Part I, this article is the chronological and thematic starting point for my dissertation. Different versions have been presented at various conferences, at *ALT 5*, Padang, Indonesia, and the conference *Approaches to Complexity in Language*, Helsinki, Finland, 2005; at the *Leipzig Students' Conference in Linguistics*, Leipzig, 2006; and at the 40th annual meeting of the *Societas Linguistica Europaea*, Joensuu, Finland, 2007. Article 1 outlines my general approach to complexity trade-offs, with a focus on the interaction of morphosyntactic coding strategies in core argument marking. As this is the oldest of the three articles in the dissertation, it also contains errors and deficiencies that reflect my knowledge at the time of writing it. Below I discuss four points on which my thinking has since changed.

1. The purpose of the article was to study whether complexity trade-offs exist in core argument marking and perhaps to evaluate whether the results would provide evidence for or against the equi-complexity hypothesis. At the time of writing, I was somewhat ambivalent as to whether studying the correlation of a few variables only would enable an evaluation of the validity of the equi-complexity hypothesis (Article 1, pp. 68, 85). As argued in Chapter 2, I now consider the equi-complexity hypothesis virtually impossible to falsify (when understood as a potential statistical universal). Regardless of this change of view, the article provides interesting results about complexity trade-offs between rigid word order and dependent marking as well as about the absence of complexity trade-offs between head marking and the two other strategies.

2. I no longer consider functional load a very promising complexity metric. The idea grew from the discussion that followed my presentation at the conference *Approaches to Complexity in Language* in 2005. A more analytical metric is to contrast the presence vs. the absence of a coding strategy, which translates directly into zero vs. non-zero complexity. This approach was used already in Article 1, but I would now emphasize using it over functional load, especially as this would enable conflating or removing the marginal and extensive functional loads, which were especially problematic to analyze.

3. At the time of writing Article 1, I was not aware of sources that enabled a later analysis of the sample languages with greater accuracy. This especially concerns the analysis of rigid order in a few languages. The biggest changes in the data analysis were the following (the source is given in parentheses):

- Nuuchahnulth was analyzed as having no rigid word order, but in light of Wojdak (2005), it uses rigid order in limited contexts to avoid ambiguity.
- Pirahã was analyzed as having no head marking (based on Everett 1986), but according to Everett (1987), both A and P can be head-marked.
- Warao was analyzed as having no dependent marking; however, the dative case can sometimes mark direct objects as well (Herrmann 2004; also Osborn 1966).
- Yelî Dnye was analyzed as having rigid word order, although I was not very confident about this analysis (see p. 84; nt. 8). In light of Levinson (2006, 2007), Yelî Dnye is now analyzed as having no rigid order.

In addition, I originally analyzed Slave as having an extensive rigid order because when the canonical SOV was changed to OSV, the object was marked on the verb (p. 75-76). However, that marking is more properly analyzed as a resumptive pronoun. This means that Slave uses rigid order in all relevant contexts, but also that the example no longer illustrates my point on morphologically marked word order. A more illuminating example comes from Trumai (Guirardello 1999), where the particle *ke* occurs clause-finally whenever the object is moved away from its canonical preverbal position (10).

Trumai (Isolate; Guirardello 1999: 57)

- (10) a. *Huch karakarako a yi disi hai-ts.*
 two chicken DU yi kill 1-ERG
 ‘I killed two chickens.’
- b. *Di heroehen wan yi ka_in hai_ts umu-ktsi ke.*
 woman beautiful PL yi FOC/TNS 1-ERG bring-DIR ke
 ‘I brought beautiful women.’

It was also a bit problematical to analyze the head marking of P in languages in which the marking was done with zero forms (e.g., in Maricopa). To determine whether the zero forms marked the P uncontroversially, it was essential to determine whether languages distinguished intransitive verbs from transitives on a lexical basis. In many instances, not enough data were found on this issue, which led to the assumption that the distinction was not a lexical one. This resulted in not classifying zero forms in a few languages as instances of the head marking of P (e.g., Maricopa, Southern Sierra Miwok). Although this assumption may be erroneous, analyzing these languages differently has no bearing on the main results, other than to change the correlation between the head marking of A and P into a significant one (see Table 4; in Article 1 this correlation was non-significant with Kendall's tau, but significant with chi-square). Overall, I consider the analysis of head marking in this article a rather weak point in my dissertation.

4. Since the time of writing Article 1, I have learned much more about statistical tests. For the kinds of data in this article, I would no longer use chi-square, but would test the correlation with Kendall's tau (or perhaps Somers' d; Somers 1962) and estimate the effect of areas with multiple logistic regression as in Article 3.

These issues suggest that it is worthwhile redoing the correlation tests to determine whether the re-analysis of the data affected the statistical tests. For these tests, I classified the coding strategies simply by noting their presence or absence. The first data-column from the left in Table 4 provides Kendall's tau values as in Article 1 (p. 81), while the right-most column provides the correlation coefficients for the reanalyzed data. As the results suggest, Kendall's tau correlation coefficient between rigid order and dependent marking was strengthened from -0.30 to -0.45, and the result also became more significant (from $p = 0.02$ to $p = 0.002$). In a similar way, the correlation between rigid order and morphological marking and that between the head marking of A and the head marking of P was strengthened, the latter so much so that it is now significant ($p = 0.023$).

The effect of areality was further tested in the light of this new data by using multiple logistic regression in a way similar to Article 3. Although the sample was picked at random, it is important to check the effect of areas, because sampling in itself

may not be a sufficient measure for controlling the effect of areas. The only pair of variables tested was the correlation between rigid order and dependent marking. Both were coded binomially with values “yes” and “no,” corresponding to the presence vs. absence of the strategy, respectively. Area was coded multinomially by using the three-way areal breakdown of Nichols (1992), who divides the world into the Old World, the Pacific, and the New World (with the Old World chosen as the baseline). The p-values were deduced by Monte Carlo randomized permutation (using 10,000 permutations). According to the results, the interaction between rigid order and area was non-significant (LR = 1.5, df = 2, p = 0.26), as was the main effect of area (LR = 1.5, df = 2, p = 0.32). However, the main effect of rigid order was significant (LR = 8.6, df = 1, p = 0.0027), and the odds for rigid order was 0.32; this means that dependent marking is more than three times less likely to occur in languages with rigid order than in those without rigid order.

Thus, despite the shortcomings in the published article, the correlation between rigid order and dependent marking remains significant in light of the reanalyzed data; in fact, it becomes even stronger than in the original article. The shortcomings thus do not cast doubt on the validity of the correlation. The results of the logistic regression analysis also provide evidence that the correlation between rigid order and dependent

Pair of variables	Correlation coefficient	
	Original data	Reanalyzed data
WO–DM	-0.302*	-0.446**
WO–HM_A	-0.035	0.000
WO–HM_P	-0.038	-0.160
WO–HM_N	-0.005	-0.080
WO–M	-0.249*	-0.364*
DM–HM_A	0.066	-0.020
DM–HM_P	-0.094	-0.088
DM–HM_N	-0.114	-0.156
HM_A–HM_P	0.201	0.320*

Table 4. Correlation coefficients for the original and reanalyzed data. (* = significant at level $p < 0.05$, ** = significant at level $p < 0.01$).

marking was independent of areas. These issues suggest that the correlation between the two variables is a valid cross-linguistic trend, but to establish it as a dynamic universal, family-internal distributions of these variables should be studied as well.

ARTICLE 2:

Complexity in core argument marking and population size

Commentary on Article 2

Article 2 grew out of studying geographical and sociocultural patterns in the distribution of complexity in the data for Article 1. The preliminary results were presented at the research seminar of general linguistics at the University of Helsinki and at the workshop *Language complexity as an evolving variable*, held in Leipzig, Germany, in 2007. Just as this article shares much of the data with Article 1, so it also shares many of the same problems. Four issues are discussed below that are related to the results. Before that, note that the references for the original chapter were listed collectively in the end of the volume and not after the chapter; these references were added right after Article 2 in the format of the originally submitted manuscript.

1. Some measures were taken to overcome problems with the data in Article 2, namely, replacing extinct languages plus a few for which the data were insufficient. This meant replacing Berbice Dutch Creole with Namia, Diyari with Pitjantjatjara, Korku with Semelai, Osage with Lakhota, Warndarang with Alawa, and Yelî Dnye with Iau. The reason for replacing Berbice Dutch Creole was that, unlike Article 1, my interest here was not in including any creoles in the sample. The reason for replacing Korku was that the sampling procedure required five languages to be sampled from Southeast Asia-Oceania (Article 1, p. 69), but Korku is spoken in India, thus, outside this area. The reasons for replacing the remaining languages had to do either with data availability or the fact that the language had already become extinct.

2. A few problems still remained in the data analyses, as follows:

- Thai was analyzed as using rigid order in all contexts, but according to Diller (1993), it allows both SOV and OSV orders. Thus, Thai is now analyzed as violating distinctiveness, the only very large language to do so.
- Babungu was analyzed as using rigid order unlike in Article 1. Based on Schaub (1985), it uses both SVO and OVS word orders, but it was difficult to assess the criteria for their usage, and in this article, I departed from the earlier analysis following Schaub (1985: 139-140). This does not affect the analysis of Babungu as violating distinctiveness, as rigid order is not used in all contexts. For the

purpose of these articles, it might have been better to have used another Bantoid language for which data on word order would have been more readily available.

- Lakhota (van Valin 1977) was analyzed as having no rigid order, but a closer inspection of the data revealed that rigid order was used, for instance, when both arguments were animate. As a result, I now analyze Lakhota as violating economy.
- Pirahã was analyzed as having no head marking, but as discussed in the Commentary to Article 1, it has limited head marking for both arguments, and so it is now analyzed as violating economy.

3. At the time of writing Article 2, it did not occur to me that the correlation between complexity and population size could have been studied with rank correlation and that the effect of areas could have been studied with logistic regression instead of testing the correlation separately in different areas. For this reason, the re-analyzed data were subjected to two more statistical tests, rank correlation and logistic regression.

For testing the correlation between complexity and population size, I used Somers' d rank correlation (Somers 1962). It is an asymmetric measure, penalizing for ties only in the predictor variable. Somers' d enables a focus on the extent to which the order of events in the independent variable, here population size, can predict the order of events in the dependent variable, here complexity; it leaves out the implausible effect of complexity on population size, unlike Kendall's tau, for instance. Complexity was coded as in Article 2, but the complexity values were now interpreted on an ordinal scale. Population size was coded binomially, as in Article 2, but the only threshold value studied here was 16,000 speakers. According to the result, complexity correlated negatively and in a very significant way with population size ($d = -0.603$; $p < 0.0001$). This confirms the original result and also shows more clearly the direction of correlation.

Next, logistic regression was used to estimate whether the correlation was independent of areas. In Article 2, the correlation did not seem independent of areas when languages violating economy were included. All of the data were used here, and complexity was coded as above for the rank correlation test. Area was coded

multinomially, using the three-way areal breakdown of Nichols (1992), as in the Commentary to Article 1. According to the results, the interaction between population size and area had no effect on complexity ($LR = 1.2$; $df = 2$; $p = 0.43$); this means that the effect of population size does not depend on areas, contrary to the conclusion in Article 2. The main effect of area was also non-significant ($LR = 4.5$; $df = 2$; $p = 0.12$), but the effect of population size was significant ($LR = 17.9$; $df = 1$; $p = 0.0001$). The odds for population size was 12.6, which means that languages with 16,000 speakers or fewer were about 13 times more likely to violate the principle of one-meaning-one-form than those with more than 16,000 speakers. These results provide further evidence for a cross-linguistic trend between complexity and population size.

4. The concept of a threshold size was crucial for testing the correlation between complexity and population size in Article 2. It is possible that there is an upper limit for population size that a community with a tight network structure could still have (Article 2, p. 138). The idea of such a threshold is that the effect of population size on language structure may become irrelevant after a certain limit: for one thing, it is unclear what difference it makes to language structure if the population size is, say, ten million rather than five million. Nevertheless, despite its apparent improbability, such differences may be relevant for language structure, as suggested by Gil (2009). For this reason, the statistical tests reported above were also conducted so that population size was coded by taking its natural logarithm, as in Lupyán and Dale (2010).

When testing the data with Somers' d , complexity correlated negatively with $\log(\text{population size})$ ($d = -0.565$; $p < 0.0001$), although not as strongly as when using a threshold for population size. When modeling the data with logistic regression, the interaction between population size and area had a significant effect on the distribution of complexity ($LR = 8.5$; $df = 2$; $p = 0.028$). This means that the effect of population size on complexity depends on area and is not universal. As was already shown in Article 2, this dependence was mostly affected by large languages in the Old World violating economy. When languages violating economy were removed, the interaction between population size and area became non-significant ($LR = 1.1$; $df = 2$; $p = 0.19$). The main effect of area was also non-significant ($LR = 0.7$; $df = 2$; $p = 0.68$), but that of population size was significant ($LR = 6.1$; $df = 1$; $p = 0.0096$).

What these results indicate is that when population size was coded by using a threshold value, there was a cross-linguistic trend between complexity and population size. On the other hand, when a natural logarithm was taken from the population size, the correlation between complexity and population size depended on areas, that is, it was not a cross-linguistic trend, an issue observed already in Article 2. When languages violating economy were removed, the correlation was a cross-linguistic trend, an issue noticed also in Article 2. Applying these new tests shows more clearly the direction of correlation and makes the original results more robust by better controlling the effect of areas. They also suggest that using a threshold value may not be necessary (*pace* Pericliev 2004) and that taking a logarithm of population size, as in Lupyan and Dale (2010), shows the trends more clearly and uses the coded information more thoroughly.

ARTICLE 3:
Word order in zero-marking languages

Commentary on Article 3

The idea for Article 3 emerged from a simple test using the *WALS* data in cooperation with Matti Miestamo in early 2005. Earlier versions of this paper have been presented at several meetings (e.g., at the conference *Rara and Rarissima*, Leipzig, Germany, 2006, at the Department of General Linguistics, Helsinki, Finland, 2007, at the conference *Language, Communication, and Cognition*, Brighton, England, 2008, and at *ALT8*, Berkeley, California, 2009). Until March 2008, I had tested the correlation between zero marking and SVO word order only with chi-square and Fisher's exact test. However, the biased distribution of zero marking caused continual difficulties, and I was unhappy with the reliability of the results at that time. A breakthrough came after listening to Balthasar Bickel's lecture at the Department of General Linguistics at the University of Helsinki in March 2008. In that lecture, he laid out the basics of evaluating universals with multiple logistic regression. I immediately saw how regression analysis could provide the missing key for the statistical analysis of my data – and so it did. This article, owing to its breadth of data coverage and depth of statistical analysis, forms the most important contribution of my dissertation to quantitative typology.

As mentioned in Article 3, logistic regression has been and is being rarely used in cross-linguistic studies. Although logistic regression is a very promising method for the evaluation of language universals, its application to typological data is not without problems. This means that currently we do not really know what the guidelines are for an acceptable regression model in typology. For example, my models might be potentially criticized as over-fitting, owing to the low number of events per variable (cf. Peduzzi et al. 1996). This issue was resolved, at least partly, by using penalized maximum likelihood ratio. The main results were further double-checked with exact logistic regression (Zamar, McNeney, and Graham 2007), which has been designed especially for data with a low number of events per variable. Since the results converged with those produced by penalized maximum likelihood ratio, there should be no problem accepting my results, despite the data limitations. An alternative method for testing sparse and biased data is random forest (Breiman 2001), but re-evaluating the results with this method has to be postponed to future research. As our experience

increases in applying regression methods to typological data, early attempts, such as Article 3, may require revision, but in light of my results, I do not expect possible revisions to change the conclusions.

In the following, I discuss briefly an issue related to regression analysis in Article 3, which I find somewhat problematic, but which was not touched on in Article 3. The issue has to do with using randomized permutation for testing the statistical significance of a penalized regression model. The script used for that purpose (the R function `rnd.lr.test.lrm()`) assigns the same penalty to all permuted datasets, namely, the ideal penalty of the original model.¹⁷ This means that the R function does not try to find the ideal penalty for each permuted dataset, but instead uses a non-ideal penalty. This may be necessary, since for many permuted datasets the ideal penalty would be infinitely large. Yet in some permuted datasets this may lead to penalizing the coefficients too much, to the extent that the coefficient becomes erroneously smaller than the observed coefficient. I have not performed systematic tests on this question, but based on my observations this situation should not be a serious problem for the method; however, it would be important to assess its effect in a future study. For now, I would suggest double-checking the borderline results by using exact logistic regression (Zamar, McNeney, and Graham 2007) despite its rather conservative output as argued by Brazzale and Davison (2008).

17 This R-function is available at <http://www.uni-leipzig.de/~autotyp/rnd.lr.test.r>.

